

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 16:47:52 ; Search time 40 seconds
(without alignments)

2877.287 Million cell updates/sec

Title: US-09-824-735-2

Perfect score: 2293

Sequence: 1 MTKMRVGVGVGRTIGEG.....IIWRATEGIPKSEILRTIF 446

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23.*

- 1: sp_archea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriopl.*
- 17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	2293	100.0	446	10	Q9LDI3	Q9ldi3 arabidopsis
2	1884.5	82.2	400	10	Q9LKR2	Q9lkr2 arabidopsis
3	1484.5	64.7	445	10	Q9STV4	Q9stv4 arabidopsis
4	1381.5	60.2	480	10	Q9C753	Q9c753 arabidopsis
5	1376.5	60.0	482	10	Q9CVN3	Q9cvn3 arabidopsis
6	1333	58.1	440	10	Q24342	Q24342 sorghum bic
7	1314	57.3	441	10	Q8RWU8	Q8rwu8 arabidopsis
8	1313	57.3	440	10	Q24343	Q24343 sorghum bic
9	1262.5	55.1	373	10	Q8GSF8	Q8gsf8 oryza sativ
10	1251.5	54.6	447	10	Q94F03	Q94f03 arabidopsis
11	1246.5	54.4	449	10	Q9C554	Q9c554 arabidopsis
12	1243.5	54.2	451	10	Q9C5P5	Q9c5p5 arabidopsis
13	1241.5	54.1	542	10	Q8SV9	Q8sv9 oryza sativ
14	1234.5	53.8	453	10	Q9MAM1	Q9mam1 arabidopsis
15	1202.5	52.4	445	10	Q8LIQ4	Q8liq4 oryza sativ
16	1149	50.1	382	10	Q9ZVD9	Q9zvd9 arabidopsis

17	1145	49.9	375	10	Q9LKC9	Q9lkc9 arabidopsis
18	1090	47.5	461	10	Q9LWM4	Q9lwm4 oryza sativ
19	1043	45.5	438	10	Q8LK24	Q8lk24 glycine max
20	1040	45.4	441	10	Q6S554	Q6s554 arabidopsis
21	1039.5	45.3	439	10	Q9FU54	Q9fu54 arabidopsis
22	1032.5	45.0	441	10	Q8W2D7	Q8w2d7 brassica na
23	1032.5	45.0	502	10	Q8LRC0	Q8lrc0 oryza sativ
24	1031.5	45.0	461	10	Q9LGV5	Q9lgv5 oryza sativ
25	1021	44.5	487	10	Q94A54	Q94a54 arabidopsis
26	1018.5	44.4	488	10	Q8W1D5	Q8w1d5 arabidopsis
27	1014.5	44.2	432	10	Q84C40	Q84c40 arabidopsis
28	1012	44.1	453	10	Q8H2C2	Q8h2c2 persea amer
29	1009.5	44.0	435	10	Q8H0X3	Q8h0x3 arabidopsis
30	1009.5	44.0	445	10	Q9LEU7	Q9leu7 arabidopsis
31	1006.5	43.9	456	10	Q9LYQ8	Q9lyq8 arabidopsis
32	1005	43.8	444	10	Q8RWC9	Q8rwc9 arabidopsis
33	1003	43.7	444	10	Q9FUK2	Q9fuk2 arabidopsis
34	1001	43.7	462	10	Q9XFJ3	Q9xfj3 mesembryant
35	999.5	43.6	421	10	Q9SX61	Q9sx61 arabidopsis
36	998.5	43.5	456	10	Q9LKD0	Q9lkd0 arabidopsis
37	996	43.4	480	10	Q9LUP6	Q9lup6 arabidopsis
38	986	43.0	464	10	Q9LVL3	Q9lvl3 arabidopsis
39	985.5	43.0	508	10	Q9SLZ6	Q9slz6 oryza sativ
40	984.5	42.9	454	10	Q9LNH7	Q9lnh7 arabidopsis
41	984	42.9	483	10	Q9FU55	Q9fu55 arabidopsis
42	983.5	42.9	489	10	Q9SN43	Q9sn43 arabidopsis
43	983	42.8	518	10	Q9SBG2	Q9sbg2 zea mays (m
44	982.5	42.8	479	10	Q9C562	Q9c562 arabidopsis
45	980	42.7	474	10	Q94DT7	Q94dt7 oryza sativ

ALIGNMENTS

RESULT 1

Q9LDI3	
ID	Q9LDI3 PRELIMINARY; PRT; 446 AA.
AC	Q9LDI3;
DT	01-OCT-2000 (TREMELrel. 15, Created)
DT	01-OCT-2000 (TREMELrel. 15, Last sequence update)
DE	01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE	Serine/threonine protein kinase SOS2 (CBL-interacting protein kinase
DE	24).
GN	SOS2 OR AT5G35410.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi
OX	NCBI_TaxID=3702;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=COLUMBIA;
RA	Kaneko T., Katon T., Asanizu E., Sato S., Nakamura Y., Kotani H.,
RA	Tabata S.;
RT	"structural analysis of Arabidopsis thaliana chromosome 5. XI.;"
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=cv. Columbia;
RX	MEDLINE=2020704; PubMed=10725382;
RA	Liu J., Ishitani M., Halter U., Kim C.S., Zhu J.K.;
RT	"The Arabidopsis thaliana SOS2 gene encodes a protein kinase that is
RT	required for salt tolerance.;"
RL	Proc. Natl. Acad. Sci. U.S.A. 97:3730-3734(2000).
RN	[3]
RP	SEQUENCE FROM N.A.
RA	Weinl S., Albrecht V., Kudla J.;
RT	"Molecular characterization of the CIPK gene family from Arabidopsis
RT	thaliana.;"
RL	Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN	[4]
RP	SEQUENCE FROM N.A.
RA	Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,

RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AB025611; BAA98146.1; -;
 DR EMBL; AF237670; AAF62923.1; -;
 DR EMBL; AF395081; AAK72257.1; -;
 DR EMBL; AY099621; AAM20472.1; -;
 DR EMBL; BT002138; AAN72149.1; -;
 DR HSP; Q63450; 1A06.
 DR InterPro; IPR000401; NAF dom.
 DR InterPro; IPR000719; Prot_kinase.
 DR Pfam; PF03822; NAF; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 446 AA; 50634 MW; 99DF3A778E1093D CRC64;
 Query Match 100.0%; Score 2293; DB 10; Length 446;
 Best Local Similarity 100.0%; Pred. No. 1.7e-183;
 Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTKMRRVKGVEVGTIGEGTFAKVFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREI 60
 DB 1 MTKMRRVKGVEVGTIGEGTFAKVFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREI 60
 QY 61 SIMKIVRHPNIVRLVLEFVTLASPSKIYVLEFVTGGELFDRIVHKGRLSESRKVFQQLVD 120
 DB 61 SIMKIVRHPNIVRLVLEFVTLASPSKIYVLEFVTGGELFDRIVHKGRLSESRKVFQQLVD 120
 QY 121 AVAHCKGVYHRDLKPNLLDNTGNNLKVSDFGLSALPQEGVELLTTCGTPNYVAPEV 180
 DB 121 AVAHCKGVYHRDLKPNLLDNTGNNLKVSDFGLSALPQEGVELLTTCGTPNYVAPEV 180
 QY 181 LSGQYDGSAAIDKSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFAEVKFL 240
 DB 181 LSGQYDGSAAIDKSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFAEVKFL 240
 QY 241 IHRILDPNPKTRIQIQGKKDPWFLNYPVIRAREEEVNLDLIRAVFDGIEGSYVAENV 300
 DB 241 IHRILDPNPKTRIQIQGKKDPWFLNYPVIRAREEEVNLDLIRAVFDGIEGSYVAENV 300
 QY 301 ERNDEGPIIMNAFEMITISQGLNLSALPDRQDFVKQTRFVSRRSEIIANIEAVNS 360
 DB 301 ERNDEGPIIMNAFEMITISQGLNLSALPDRQDFVKQTRFVSRRSEIIANIEAVNS 360
 QY 361 MGFKSHTRNFTRFGLSSIKAGQAVVIEIYEVAPSLFMVDVRKAAGETLEYHKFYKKL 420
 DB 361 MGFKSHTRNFTRFGLSSIKAGQAVVIEIYEVAPSLFMVDVRKAAGETLEYHKFYKKL 420
 QY 421 CSKLENIWRATEGIPKSEILRTITTF 446
 DB 421 CSKLENIWRATEGIPKSEILRTITTF 446

RESULT 2
 Q9LKR2

ID Q9LKR2 PRELIMINARY; PRT; 400 AA.
 AC Q9LKR2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE F6113.1 protein (fragment).
 GN F6113.1
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxId=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA WashU;
 RT "The A. thaliana Genome Sequencing Project.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Wilson R.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AF262044; AAF67384.1; -;
 DR HSP; Q63450; 1A06.
 DR InterPro; IPR000401; NAF dom.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR Pfam; PF03822; NAF; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 FT NON_TER 400
 SQ SEQUENCE 400 AA; 45623 MW; 6908A8A695E12BAF CRC64;
 Query Match 82.2%; Score 1884.5; DB 10; Length 400;
 Best Local Similarity 93.2%; Pred. No. 2.5e-149;
 Matches 371; Conservative 0; Mismatches 0; Indels 27; Gaps 2;
 QY 1 MTKMRRVKGVEVGTIGEGTFAKVFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREI 60
 DB 1 MTKMRRVKGVEVGTIGEGTFAKVFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREI 60
 QY 61 SIMKIVRHPNIVRLY-----EVLASPSKIYVLEFVTGGELFDRI----- 100
 DB 61 SIMKIVRHPNIVRLYEVCLFVSMHLRNFISEVLASPSKIYVLEFVTGGELFDRIVRNFH 120
 QY 101 -----VHKGRLSESRKVFQQLVDVAHCHCKGVYHRDLKPNLLDNTGNNLKVSDF 153
 DB 121 TCQRPLVHKGRLSESRKVFQQLVDVAHCHCKGVYHRDLKPNLLDNTGNNLKVSDF 180
 QY 154 GLSALPQEGVELLTTCGTPNYVAPEVLSGGYDGSAAIDKSCGVILFVILAGYLPFSET 213
 DB 181 GLSALPQEGVELLTTCGTPNYVAPEVLSGGYDGSAAIDKSCGVILFVILAGYLPFSET 240
 QY 214 DLPLGYRKINAAEFSCPPWFAEVKFLIHRILDPNPKTRIQIQGKKDPWFLNYPVIRA 273
 DB 241 DLPLGYRKINAAEFSCPPWFAEVKFLIHRILDPNPKTRIQIQGKKDPWFLNYPVIRA 300
 QY 274 REEEVNLDLIRAVFDGIEGSYVAENVENDEGPIIMNAFEMITISQGLNLSALPDRQD 333
 DB 301 REEEVNLDLIRAVFDGIEGSYVAENVENDEGPIIMNAFEMITISQGLNLSALPDRQD 360
 QY 334 FVKRQTRFVSRRSEIIANIEAVNSMGFKSHTRNFK 371
 DB 361 FVKRQTRFVSRRSEIIANIEAVNSMGFKSHTRNFK 398

```

RESULT 3
Q9STV4 PRELIMINARY; PRT; 445 AA.
AC Q9STV4; Q9C5S5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 23, Last annotation update)
DE Serine/threonine kinase-like protein (CBL-interacting protein kinase
DE 8).
GN T22a6.230 OR A74G24400 OR CIPK8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Zimmermann W., Gruenisen A., Wambutt R., Bancroft I.,
RA Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Zimmermann W., Gruenisen A., Wambutt R., Kalicki J., Wohldmann P.,
RA Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Zimmermann W., Gruenisen A., Wambutt R., Kalicki J., Wohldmann P.,
RA Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA PubMed=11230129;
RX Albrecht V., Ritz O., Linder S., Harter K., Kudla J.;
RA "The NAF domain defines a novel protein-protein interaction module
RT conserved in Ca(2+)-regulated kinases.";
RL EMBL J. 2011051-1063 (2001).
RN [6]
RP SEQUENCE FROM N.A.
RA Kudla J., Albrecht V.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Huan V.W., Lee J.M.,
RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,
RA Ecker J.R., Theologis A.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RA "Arabidopsis Open Reading Frame (ORF) Clones.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AL078637; CAB45075.1; -
DR EMBL; AL161561; CAB79350.1; -
DR EMBL; AF290193; AAK16683.2; -
DR EMBL; BT000958; AAN41358.1; -
DR HSP; Q63450; 1A06.
DR InterPro; IPR004041; NAF dom.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF03822; NAF; 1.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 445 AA; 50428 MW; 46636F20F327E709 CRC64;

Best Local Similarity 62.3%; Pred. No. 9.2e-116;
Matches 273; Conservative 84; Mismatches 80; Indels 1; Gaps 1;
QY 5 MERVGVYVGRITIGTGTAKVKFAENTDTGDNVAIKMAKSTILKNRVWDQIKRISIMK 64
DB 3 VRKVGYELGRITIGTGTAKVKFAONTETGESVAMKIVDRSTLIIRKRWQDKRISIMK 62
QY 65 IVRHPNIVRLVEVLASPSKIYIVLEFVTGGELFDRIVHKGRLSESESRKYFOOLVDAVAH 124
DB 63 LVRHPCVVRVLEVLASRTKIYIVLEFVTGGELFDKIVENGRLESESEARKYFHLIDVDY 122
QY 125 CHCKGVYHDKLPENLLDWTGNLKVSPGLSALPDQGVELLRTTCGTFYVAVPEVLSGQ 184
DB 123 CHSKGVYHDKLPENLLDWTGNLKVSPGLSALPDQGVELLRTTCGTFYVAVPEVLSHK 182
QY 185 GYDGSAAADIWSCGVILFVILAGYLPFSETDPLGVRKINAAEFSCPPWFSAVRKFLIHR 244
DB 183 GYNGAVADIWSCGVILYVLMAGYLPFDEMDLPTLYSKIDKAEFSCPSYFALGAKSLINRI 242
QY 245 LDPNPKTRIQIGIKKDPWFRINYPVIRAREEEVNLDDIRAVFDGIEGYSYVAENVERND 304
DB 243 LDPNPETRITIAEIRKDEWFLKDYTPVQLIDYEHVNLDDVYAAFDDPEQTVADQGT-R 301
QY 305 EGPLMMNAFEMITLSOGLNLGALFDRQDFVKQTRFVSRREPSELIANIEAVANMGEFK 364
DB 302 TGPLTLNADFLLISQGLNLATLFDGRKDSMKHQTRFISHKPPANVYLSMEVVSQSGFK 361
QY 365 SHTRNFKTRLEGLSISIKAGQLAVWIEYEVAPSLFMVDVRAAGETLBYHKFYKLSCKL 424
DB 362 THIRYKRVGSLGANKTSHPFVILEVFKVAPSLMVDIQNADGAEYLFYKTFCSKL 421
QY 425 ENIWRATEGIPKSEILR 442
DB 422 DDIIWKPPDASMRNRVTK 439

RESULT 4
Q9C753 PRELIMINARY; PRT; 480 AA.
AC Q9C753;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Serine/threonine kinase, putative.
GN F12P21.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Columbia;
RC MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.L.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uteback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820 (2000).

```

```
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AC073506; AAG50566.1; -.
DR HSP; Q63450; 1A06.
DR InterPro; IPR004041; NAF dom.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF03822; NAF; 1.
DR Pfam; PF0069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 480 AA; 53259 MW; 89E09C800ABF925E CRC64;

Query Match 60.2%; Score 1381.5; DB 10; Length 480;
Best Local Similarity 60.2%; Pred.No. 4.4e-107;
Matches 260; Conservative 78; Mismatches 85; Indels 9; Gaps 4;

QY 7 RVGKYEVGRTIGEGTFAKVFARNTDTGDNVAIKMAKSTILKNRMVDPQIKREISIMKIV 66
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 27 RVGKYELGRTIGEGTFAKVFARNVNGDNVAIKVLDKEKVLKNKMIQIKREISTMKLI 86
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 67 RHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGLESESRKYFOQLVDVAHCH 126
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 87 KHPNIVRMFEVNASKTIYFVLEFVTGGELFDKISSNGLKDEARKYFQQLINAVDYCH 146
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 127 CKGVVHRDLKPNLLDITNGNLKVSDFGLSALPQEGVELLRTTCGPNVVAPEVLSGGY 186
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 147 SGVYHRDLKPNLLDITNGNLKVSDFGLSALPQEGVELLRTTCGPNVVAPEVLSGGY 206
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 187 DGSADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVEKFLIHR 246
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 207 DGAKADLWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVEKFLIHR 266
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 247 PNPKTIRIQIGIKKDPWFRLNVPFIRAREEVEVNDLDIRAVFDGIEGVSVAEN--VER 304
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 267 PNPATRIITFAVIEENEFKGYKAPKF-ENADVSLDDVDAIFD---DSGESKNLVVER 322
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 305 EG---PLMNAFEMITLSQGLNLKVSDFGLSALPQEGVELLRTTCGPNVVAPEVLSGGY 361
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 323 EGLKTPVTWNAFELISTSGNLKVSDFGLSALPQEGVELLRTTCGPNVVAPEVLSGGY 382
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 362 GKPSHTRPKTRLEGSSIKAGQLAVIYIYVAPSLFMDVVRKAAGETLEYHFKYKLC 421
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 383 GPDVKTNNYKMKLTGKSGKGLAVATEVFPVAPSLYVMEMRKSGGDTLEPHKFKYKLT 442
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 422 SKLENIWRATE 433
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 443 TGLKDLVWKTID 454
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 5
Q93VD3 PRELIMINARY; PRT; 482 AA.
AC Q93VD3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Atlg30270/F12P21.6 (CBL-interacting protein kinase 23).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Shinn P., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
```

```
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Weinl S., Albrecht V., Kudla J.;
RT "Molecular characterization of the CIPK gene family from Arabidopsis
thaliana.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AY056419; AAL08275.1; -.
DR EMBL; AY035226; AAK61494.1; -.
DR EMBL; AY090322; AAL90983.1; -.
DR HSP; P24941; 1BUH.
DR InterPro; IPR004041; NAF dom.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF03822; NAF; 1.
DR Pfam; PF0069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 482 AA; 53514 MW; CAA1B526B955F5F2 CRC64;

Query Match 60.0%; Score 1376.5; DB 10; Length 482;
Best Local Similarity 60.1%; Pred.No. 1.2e-106;
Matches 261; Conservative 78; Mismatches 84; Indels 11; Gaps 5;

QY 7 RVGKYEVGRTIGEGTFAKVFARNTDTGDNVAIKMAKSTILKNRMVDPQIKREISIMKIV 66
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 27 RVGKYELGRTIGEGTFAKVFARNVNGDNVAIKVLDKEKVLKNKMIQIKREISTMKLI 86
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 67 RHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGLESESRKYFOQLVDVAHCH 126
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 87 KHPNIVRMFEVNASKTIYFVLEFVTGGELFDKISSNGLKDEARKYFQQLINAVDYCH 146
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 127 CKGVVHRDLKPNLLDITNGNLKVSDFGLSALPQEGVELLRTTCGPNVVAPEVLSGGY 184
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 147 SGVYHRDLKPNLLDITNGNLKVSDFGLSALPQEGVELLRTTCGPNVVAPEVLSGGY 206
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 185 GYDGAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVEKFLIHR 244
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 207 GYDGAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVEKFLIHR 266
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 245 LDPNPKTRIQIGIKKDPWFRLNVPFIRAREEVEVNDLDIRAVFDGIEGVSVAEN--VER 302
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 267 LDPNPKTRITFAVIEENEFKGYKAPKF-ENADVSLDDVDAIFD---DSGESKNLVVER 322
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 303 NDEG---PLMNAFEMITLSQGLNLKVSDFGLSALPQEGVELLRTTCGPNVVAPEVLSGGY 359
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 323 REGLKTPVTWNAFELISTSGNLKVSDFGLSALPQEGVELLRTTCGPNVVAPEVLSGGY 382
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 360 SMGFKSHTRPKTRLEGSSIKAGQLAVIYIYVAPSLFMDVVRKAAGETLEYHFKYKLC 419
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 383 PMGPDVKTNNYKMKLTGKSGKGLAVATEVFPVAPSLYVMEMRKSGGDTLEPHKFKYKLC 442
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```



```

QY 420 LCSKLENIIRATE 433
Db 443 LTGLKDVWKTID 456

RESULT 6
ID O24342 PRELIMINARY; PRT; 440 AA.
AC O24342;
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Serine/threonine kinase.
GN SNFL1.
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
OX NCBI_TaxID=4558;
RN [1]_TaxID=4558;
RP SEQUENCE FROM N.A.
RC STRAIN=cv. TX 430;
RX MEDLINE=98145442; PubMed=9484448;
RA Annen F., Stockhaus J.;
RT "Characterization of a Sorghum bicolor gene family encoding putative
RT protein kinases with a high similarity to the yeast SNF1 protein
RT kinase";
RL Plant Mol. Biol. 36:529-539(1998).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; Y12464; CAAT3067.1; -.
DR HSSP; Q63450; 1A06.
DR InterPro; IPR004041; NAF dom.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF03822; NAF; 1.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKc_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 440 AA; 50458 MW; D71B8E34B4E0E73C CRC64;

Query Match
Best Local Similarity 58.1%; Score 1333; DB 10; Length 440;
Matches 258; Conservative 72; Mismatches 98; Indels 6; Gaps 3;

QY 2 TKMRVRGVKYEGRITIGETFAKVFARNTDGTGDNVAIKIMAKSTILKNRVDQIKREIS 61
Db 4 TKVRVRGVKYEGRITIGETFAKVFARNTDGTGDNVAIKIMAKSTILKNRVDQIKREIS 63
QY 62 IMKIVRHPNIVLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLSESRKYFOOLVDA 121
Db 64 TMKLIKHPNIVLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLSESRKYFOOLVDA 123
QY 122 VAHCCKGVYHDLKPENLITNGNLKVSDFGLSALPOEGVE--LLRTTCGTNNVAPE 179
Db 124 VYCHSRGVYHDLKPENLITNGNLKVSDFGLSALPOEGVE--LLRTTCGTNNVAPE 183
QY 180 VLSGGYDGSADISWCGVILFVILAGYLPFFSETDLPGLYRKINAEPFCPPWFSAEVKF 239
Db 184 VLEDQYDGDAMADLWSGCVILFVILAGYLPFFSETDLPGLYRKINAEPFCPPWFSAEVKF 243
QY 240 LIHRILDPNPKTRIQIGTKDQFVKTQTRVFSRREPSEIIANIEVAN 299
Db 244 LLTRILDPNPKTRIQIGTKDQFVKTQTRVFSRREPSEIIANIEVAN 300
QY 300 VERNDEGLPMNAFEMITLSQGLNLSALFDQDQFVKTQTRVFSRREPSEIIANIEVAN 359
Db 301 TEKKEEPVALNAFELISMSAGLNGLNFDSEQEP--KRETRTSKPPKEIVRKIEEAK 359
QY 360 SMGFKSHTRNFKTRLEGLSSIKAGQLAVVIEYVAPSLFMVDVRAAGETLEYHFKYK 419

us-09-824-735-2.rspt
Page 5

```



```

Db 7 LRVGKVEVGTIGEGTFAKVKFARNTDGNVAIKIMAKSTILKNRMVDOIKREISIMK 66
QY 65 IVRHNPVRLVEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESRKYFQQLVDAVAH 124
Db 67 LVHRPNV-----VLASRKFIILFETITGGELFDKIRHGRLENEADARRVFOQLIDGVDF 121
QY 125 CHCKGYVHRDLKPNLLDNGNLKVSDFGLSALPQEGVELLRTTCGTNYVAPEVLSGQ 184
Db 122 CHSKGYVHRDLKPNLLDNGNLKVSDFGLSALPQEGVELLRTTCGTNYVAPEVLSGQ 181
QY 185 GYDGAADTWSGCVILFVILAGYLPSETDPLGLYRKINAABSCPPWESAEVKFLIHR 244
Db 182 GYDGAADTWSGCVILFVILAGYLPSETDPLGLYRKINAABSCPPWESAEVKFLIHR 241
QY 245 LDNPETRIQIGIKKDPWFLNYPVIRAREEEVNLDITRAVFDGIEGYSVAENVERN 304
Db 242 LDNPETRIQIGIKKDPWFLNYPVIRAREEEVNLDITRAVFDGIEGYSVAENVERN 299
QY 305 EGPLMNAFEMITLSQGLNLSALFDRQDFVKKQTRFVSRREPSEIIANIEAVANSG 364
Db 300 AGPLTUNARDFLLIISQGLNLSALFDRQDFVKKQTRFVSRREPSEIIANIEAVANSG 359
QY 365 SETRNEFT 372
Db 360 THIRNYKT 367

RESULT 10
Q94F03 PRELIMINARY; PRT; 447 AA.
ID Q94F03
AC Q94F03
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Similar to wpk4 protein kinase.
GN T25K16.13 OR AtG01140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lam B., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Bath J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Bath J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AF386999; RAK6244.1; -
DR EMBL; AY093242; AAL13241.1; -
DR HSSP; P24941; 1BUH.
DR InterPro; IPR004041; NAF dom.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF03822; NAF; 1.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.

```

```

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 447 AA; 50476 MW; D6AGBF7B813760C5 CRC64;

Query Match 54.6%; Score 1251.5; DB 10; Length 447;
Best Local Similarity 54.7%; Pred. No. 3.1e-96;
Matches 233; Conservative 85; Mismatches 101; Indels 7; Caps 3;

QY 7 RVGKVEVGTIGEGTFAKVKFARNTDGNVAIKIMAKSTILKNRMVDOIKREISIMKIV 66
Db 15 RVGNVEMGTIGEGSFAKAKYAKNTVTGQAAIKILDRKVFHKVGEQKREISTMKLI 74
QY 67 RHPNIVRLVEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESRKYFQQLVDAVAHCH 126
Db 75 KHPNVVEILVEVASKTKIYIVLELVNGGELFDKIAQQGRLEDEARRVYFQQLINAVDYCH 134
QY 127 CHCKGYVHRDLKPNLLDNGNLKVSDFGLSALPQEGVE--LRTTCGTNYVAPEVLSGQ 184
Db 135 SRGVHRDLKPNLLDNGNLKVSDFGLSALPQEGVE--LRTTCGTNYVAPEVLSGQ 194
QY 185 GYDGAADTWSGCVILFVILAGYLPSETDPLGLYRKINAABSCPPWESAEVKFLIHR 244
Db 195 GYDGAADTWSGCVILFVILAGYLPSETDPLGLYRKINAABSCPPWESAEVKFLIHR 254
QY 245 LDNPETRIQIGIKKDPWFLNYPVIRAREEEVNLDITRAVFDGIEGYSVAENVERN 303
Db 255 LDNPETRIQIGIKKDPWFLNYPVIRAREEEVNLDITRAVFDGIEGYSVAENVERN 313
QY 304 DEGPLMNAFEMITLSQGLNLSALFDRQDFVKKQTRFVSRREPSEIIANIEAVANSG 363
Db 314 ---PVSMAFELISSSEFSLENLPEKQALVKETRTSQSASEIMSKMETAKPLGF 370
QY 364 KSHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMDVDRKAAAGTIEYHKFKLCSK 423
Db 371 NVKDKNYIKMKGDKSRGKQLSVATEFEVAPSLHVLRLTGTDLTFHFHAFYKNFSSG 430
QY 424 LENIIV 429
Db 431 LKDVVW 436

RESULT 11
Q9CS54 PRELIMINARY; PRT; 449 AA.
ID Q9CS54
AC Q9CS54
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CBL-interacting protein kinase 9.
GN CIPK9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21153204; PubMed=11230129;
RA Albrecht V., Ritz O., Linder S., Harter K., Kudla J.;
RT "The NAF domain defines a novel protein-protein interaction module
RL conserved in Ca(2+)-regulated kinases."
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AF295664; AKJ16684.1; -
DR HSSP; Q63450; 1A06.
DR InterPro; IPR004041; NAF dom.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF03822; NAF; 1.
DR ProDom; PD00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

```


DR PROSITE; P500011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; P500108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 542 AA; 61152 MW; 0462D0A2F4427F3 CRC64;

Query Match 54.1%; Score 1241.5; DB 10; Length 542;
Best Local Similarity 54.9%; Pred. No. 2.8e-95;
Matches 237; Conservative 85; Mismatches 103; Indels 7; Gaps 3;

QY 3 KQMRVKGVEGRTIGETFAKVFARNTDGDNNVAIKMAKSTILKNRWVDQIKREISL 62
Db :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
19 RRTTRVGRLVKTKTIGESFAKVVARTRTGDTALIKVLDRNHLRHKWVEQIKREIST 78
Db ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 63 MKIVRHNPVRLIYELASPSKIYIIVLEVTGGELFDRIYHVKGRLESESRKYFOQLVDVAV 122
Db ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
79 MKLIKHPNVQLHEVMASKSIYMLVLYVGGELFDKIIVNSGRGLGEDEARRYPHQLINAV 138
Db ::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 123 AHCHCKGVHRDLKPENLLDNGNLKVSDFGLSALPOSGVE--LLTTTCGTPTYVAPEV 180
Db :||::||::||::||::||::||::||::||::||::||::||::||::||:
139 DYCHRGVYHRDLKPENLLDSHGALKVSDFGLSAFAPQTQKEDGLLTACGTPTYVAPEV 198
Db :||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 181 LSGQGVDGAADIMWCGVTLFVILAGYLPPSETDLPGLYRKINAAEFSCPPWFSAEVKFL 240
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
199 LADKGYDGWAADVWSGCIILFVLMAGYLPFPDDPNMILYKLICAKVUSCPHWSSGAKKF 258
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 241 IHRILDNPKTIRIOIGIKKDPWFRNLNYVPPIRARBEENVLLDIRAVPDGIEGSYAENV 300
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
259 IKRILDNPCTRTITIAILEDWFKDKYPPLFEQGEDVSLDDVDAADFDCSEENLVAEKR 318
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 301 ERNDGCLPMNAFEMITISQGLNLSALPDR-QDFVKQTRFVSRRPSPSIIANIEAVAN 359
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
319 EK----PVSMMNALISSSFSSSELENLFKQAQLVKKTRFTSQRSASEINSMKEETA 374
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 360 SMGFKSHTRNFKTRLEGSLSIKAGOLAVVIETIVEVAPSLEFMDVRKAAGTLEVHKFYKK 419
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
375 PLGFNVKQNYKMKGDKSGRKGQLSVATEVFEVAPSLHHVELKRTGGDTLTFHKVCDSFY 430
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 420 LCSKLENIIRA 431
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
435 FSSELKDIIWKKS 446

RESULT 14
Q9MAM1 PRELIMINARY; PRT; 453 AA.

ID	Q9MAM1	PRELIMINARY;	PRT;	453 AA.
AC	Q9MAM1;			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	T25K16.13.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Dunn P., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,			
RA	Khan S., Kim C., Altafi H., Bei O., Chin C., Chiou J., Choi E.,			
RA	Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,			
RA	Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,			
RA	Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,			
RA	Thaveri A., Toriumi M., Vaysberg M., Yu G., Federspiel N.A.,			
RA	Theologis A., Ecker J.R.;			
RT	"Genomic sequence for Arabidopsis thaliana BAC T25K16 from chromosome			
RT	I.",			
CC	Submitted (JUN-2000) to the EMBL/GenBank/DBSJ databases.			
RC	-I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.			
DR	EMBL; ACC07323; AAF26468.1; -			
DR	HSSP; Q63450; 1A06			
DR	InterPro; IPR004041; NAF dom.			
DR	InterPro; IPR000719; Prot_kinase.			
DR	InterPro; IPR002290; Ser_thr_kinase.			

PFam; PF03822; NAF; 1.
PFam; PF00069; pkinase; 1.
ProDom; PD00001; Prot_kinase; 1.
SMART; SM00220; S_TKC_1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 453 AA; 51164 MW; 544789B24758285 CRC64;

Query Match 53.8%; Score 1234.5; DB 10; Length 453;
Best Local Similarity 53.9%; Pred. No. 8.3e-95;
Matches 233; Conservative 86; Mismatches 100; Indels 13; Gaps 5;

QY 7 RVGKYEVRTIGETFAKVFARNTDGDNNVAIKMAKSTILKNRWVDQIKREISMKIV 66
Db ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
15 RVGNVEMERTIGESFAKVAKTVTGDQAAIKILDREKVFHKWVEQIKREISMKLI 74
Db ||::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 67 RHPNIVRLIYELASPSKIYIIVLEVTGGELFDRIYHVKGRLESESRKYFOQLVDVAHCH 126
Db :||::||::||::||::||::||::||::||::||::||::||::||::||:
75 KHPNVVEIIEWASKTIYIIVLELVNGGELFDKIAQQOGLKDEEARRYFOQLINAVDYCH 134
Db ||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 127 CKGWYHRDLKPENLLDNGNLKVSDFGLSALPOSGVE--LLTTTCGTPTYVAPEVLSGQ 184
Db :||::||::||::||::||::||::||::||::||::||::||::||::||:
135 SRGVYHRDLKPENILLDANGVLKVSDFGLSAFQRVEDGLLTACGTPTYVAPEVLSDK 194
Db ||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 185 GYDGAADIMWCGVTLFVILAGYLPPSETDLPGLYRKIN-AAEFSCPPWFSAEVKFLIH 242
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
195 GYDGAADVNSCGVTLFVLMAGYLPFPDDPNMILYKRVICKAEFSCPPWFSAEKARV 254
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 243 RILDNPKTIRIOIGIKKDPWFRNLNYVIPR-ARBEENVLLDIRAVPDGIEGSYAENV 301
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
255 RILENPITRISIALLLEDWFKGKPKPSDFDQDDDETITDDVDAAFSNSKECLVTEKKE 314
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 302 RNDEGPLMNNAFEMITISQGLNLSALPDRQDFVKQTRFVSRREPSEIIANTIEAVNSM 361
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
315 K-----PVSMNALISSSFSSSELENLFKQAQLVKKTRFTSQRSASEINSMKEETA 370
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY 362 GFKSHTRNFKTRLEGSLSIKAGOLAVVIETIVEVAPSLEFMDVRKAAGTLEVHK---FY 417
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:
371 GFNVKQNYKMKGDKSGRKGQLSVATEVFEVAPSLHHVELKRTGGDTLTFHKVCDSFY 430
Db |::||::||::||::||::||::||::||::||::||::||::||::||::||:

RESULT 15
Q8LIG4 PRELIMINARY; PRT; 445 AA.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 16:47:01 ; Search time 15 Seconds
(without alignments)
1398.261 Million cell updates/sec

Title: US-09-824-735-2

Perfect score: 2293
Sequence: 1 MTKKRRVGVKVEVGTIGEG.....IIWRATEGIPKSEILRTITF 446

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	701.5	30.6	576	1 SNF1 SCHPO	Q74536 schizosacch
2	700.5	30.5	512	1 K110 ARATH	Q74536 arabidopsis
3	673	29.4	611	1 SNF1 CANGA	Q00372 candida gla
4	670	29.2	552	1 AAK2 HUMAN	P54846 homo sapien
5	665	29.0	619	1 SNF1 CANTR	Q94168 candida tro
6	663	28.9	633	1 SNF1 YEAST	P06782 saccharomyc
7	662	28.9	552	1 AAK2 RAT	Q09137 rattus norv
8	661.5	28.8	620	1 SNF1 CANAL	P52497 candida alb
9	655.5	28.6	550	1 AAK1 HUMAN	Q13131 homo sapien
10	653	28.5	548	1 AAK1 RAT	P54845 rattus norv
11	610	26.6	502	1 RK11 SECCE	Q02723 secale cere
12	598	26.1	1142	1 GIN4 YEAST	Q12263 saccharomyc
13	592.5	25.8	622	1 YNA3 CABEL	P45894 caenorhabdi
14	589.5	25.7	1518	1 KK11 YEAST	P34244 saccharomyc
15	586	25.6	752	1 MRK4 HUMAN	Q96134 homo sapien
16	576.5	25.1	776	1 SN1L RAT	Q91u5 rattus norv
17	566.5	24.7	779	1 SN1L MOUSE	Q60670 mus musculu
18	563.5	24.6	776	1 MRK3 HUMAN	P27448 homo sapien
19	561.5	24.5	774	1 MRK2 MOUSE	Q05512 mus musculu
20	561.5	24.5	786	1 SN1L HUMAN	P57059 homo sapien
21	555.5	24.2	915	1 KCC4 YEAST	P25389 saccharomyc
22	541	23.6	800	1 K1N4 YEAST	Q01919 saccharomyc
23	534.5	23.3	775	1 CDR2 SCHPO	P87050 schizosacch
24	530.5	23.1	661	1 Y537 HUMAN	O60285 homo sapien
25	520.5	22.7	408	1 ST6L XENLA	Q91819 xenopus lae
26	517.5	22.6	407	1 STK6 XENLA	Q91820 xenopus lae
27	513	22.4	714	1 HUNK HUMAN	P57058 homo sapien
28	510.5	22.3	353	1 ASK2 ARATH	P43292 arabidopsis
29	510	22.2	332	1 AA1P WHEAT	Q02866 triticum ae
30	506	22.1	476	1 CHK1 HUMAN	Q14757 homo sapien
31	504.5	22.0	714	1 HUNK MOUSE	O88866 mus musculu
32	495	21.6	593	1 CDR1 SCHPO	P07334 schizosacch
33	492	21.5	476	1 CHK1_MOUSE	O35280 mus musculu

RESULT 1

ID	SNF1 SCHPO	STANDARD;	PRT;	576 AA.
AC	O74536;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	SNF1-like protein kinase (EC 2.7.1.-).			
GN	SPCC74.03C.			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
OX	NCBI_TaxID=4896;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=972;			
RX	MEDLINE=21848401; PubMed=11859360;			
RA	Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,			
RA	Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,			
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,			
RA	Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,			
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,			
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,			
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,			
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,			
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,			
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,			
RA	Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,			
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,			
RA	Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,			
RA	Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,			
RA	Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,			
RA	Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,			
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,			
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,			
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,			
RA	Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,			
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,			
RA	Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg J.,			
RA	Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,			
RA	Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;			
RT	"The genome sequence of Schizosaccharomycetes pombe";			
RL	Nature 415:871-880(2002).			
CC	-!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.			
CC	-!- SNF1 SUBFAMILY.			
CC	-!- SIMILARITY: Contains 1 UBA domain.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			

ALIGNMENTS


```

DR EMBL; M93023; AAA32736.1; --
DR EMBL; X79707; CA56146.1; --
DR EMBL; AC008261; AAF26165.1; --
DR EMBL; X94757; CA64384.1; --
DR PIR; JCI446; JCI446.
DR HSSP; Q63450; IA06.
DR InterPro; IPR001772; Kinase Cterm.
DR InterPro; IPR000719; Prot Kinase.
DR InterPro; IPR002290; Ser Thr Kinase.
DR InterPro; IPR001245; Tyr Kinase.
DR InterPro; IPR000449; UBA_domain.
DR Pfam; PF02149; UBA; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00627; UBA; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot kinase; 1.
DR SMART; SM00220; S_TK; 1.
DR SMART; SM00165; UBA; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00030; UBA; 1.
DR Transferase; Serine/threonine-protein kinase; ATP-binding;
KW Phosphorylation.
FT DOMAIN 19 271 PROTEIN KINASE.
FT NP_BIND 252 332 UBA.
FT BINDING 25 33 ATP (BY SIMILARITY).
FT ACT_SITE 48 48 ATP (BY SIMILARITY).
FT ACT_SITE 142 142 BY SIMILARITY.
FT MOD_RES 175 175 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 512 AA; 58373 MW; 5A18655AAAS06DF CRC64;

Query Match 30.5%; Score 700.5; DB 1; Length 512;
Best Local Similarity 32.6%; Pred. No. 9.8e-43;
Matches 155; Conservative 97; Mismatches 143; Indels 79; Gaps 9;

QY 11 YEVGTICGEGFAKVEKARNYDTGDNVAIKMAKSTILKNRMVQIKKEISIMKIV 70
Db 19 YKLGTLGSGRGRVIAHAHTGHKVAIKILNRKIKMNEBEKVRREIKILFVHPH 78

QY 71 IVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLSESRKYFQQLVDAVAHCHCKGV 130
Db 79 IIRLYEVIETPTDIYLVMEYVNSGELFDYIVKGLQDEARNFFQIISGVEYCHRMV 138

QY 131 YHRDLKPENLLDNTGNLKVSDPGLSALPQEGVELLRTTCGTPNVVAPDEVLSGGYDGA 190
Db 139 YHRDLKPENLLDNTGNLKVSDPGLSALPQEGVELLRTTCGTPNVVAPDEVLSGGYDGA 197

QY 191 ADIWSGVLIFVLGALYLPFSETDLPGLYRKINAAEFSCPPWFAEVKFLIHRILDPNPK 250
Db 198 VDWSCGVILYVALLCGTLPFDENIPNLFKKIKGGIYTLPSHLSPGARDLIPMLVVDPM 257

QY 251 TRIQGIKKDWFLN---YVPI-----RAREEVNLDI-----284
Db 258 KEVTPTEIRQHPWFQAHLPYLAIVPPPTVQQAQKIDELIQEVINMGFDRNHILIESLRN 317

QY 285 RAVFDD-----IEGSVAENVVERDEGLMNAFMTITLSQGLNLGALFD--329
Db 318 FTQNDGTYYIILNDRFASGYLGAEFQETMEGTPMHPAESAIVSPVSHRLPGLMEYQ 377

QY 330 ----RQDFVKQ--TRFVSRREPSEIANI-----354
Db 378 GVGRLSQYVERKVALGLQSRAPRREIMTEVLKALQDLNVCKKTHGYNMKCRVWPNSSA 437

QY 355 -EAVANSMGFKSHTRFKRLRGLSISIKAGQLA--VVIIEYVAPSLFVVDYRKAAG 408
Db 438 DGMLSNMDHNNFYGDESSIIENEAHSVNVVKEFIOIYKTRDDKYLIDLRQVQ 493

RESULT 3
SNFL_CANGA
ID_SNF1_CANGA STANDARD; PRT; 611 AA.
AC Q00372;

```

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
GN SNF1.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]_TaxID=5478;
RP SEQUENCE FROM N.A.
RC STRAIN=NCCLS84;
RX MEDLINE=97101049; PubMed=8945576;
RA Pether R., Kwon-Chung K.J.;
RT "Disruption of the SNF1 gene abolishes trehalose utilization in the
RL pathogenic yeast Candida glabrata.";
RL Infect. Immun. 64:5269-5273(1996)
CC -!- FUNCTION: ESSENTIAL FOR RELEASE FROM GLUCOSE REPRESSION. IT
CC INTERACTS AND HAS FUNCTIONAL RELATIONSHIP TO THE REGULATORY
CC PROTEIN SNF4. COULD PHOSPHORYLATES CAT8 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE NUCLEAR MEMBRANE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC SNF1 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; L78130; AAB48642.1; --
CC HSSP; P24941; 1HCL.
CC InterPro; IPR000719; Prot kinase.
CC InterPro; IPR002290; Ser Thr kinase.
CC Pfam; PF00069; pkinase; 1.
CC ProDom; PD000001; Prot kinase; 1.
CC SMART; SM00220; S_TK; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC Transferase; Serine/threonine-protein kinase; ATP-binding;
KW Phosphorylation; Carbohydrate metabolism; Nuclear protein.
FT DOMAIN 6 17 POLY-HIS.
FT NP_BIND 39 230 PROTEIN KINASE.
FT BINDING 45 53 ATP (BY SIMILARITY).
FT ACT_SITE 68 68 ATP (BY SIMILARITY).
FT ACT_SITE 161 161 BY SIMILARITY.
FT MOD_RES 194 194 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 611 AA; 70049 MW; 89E17812A4900CD0 CRC64;

Query Match 29.4%; Score 673; DB 1; Length 611;
Best Local Similarity 38.8%; Pred. No. 1.1e-40;
Matches 142; Conservative 81; Mismatches 113; Indels 30; Gaps 8;

QY 7 RVGKYEVGRTIGEGTFAKVEKARNYDTGDNVAIKMAKSTILKNRMVQIKKEISIMKIV 66
Db 35 RVGNVQIVLTGEGSGFKVRLAYHTTGQVAKLSINKVLAKSDMQGRIDREISYRL 94

QY 67 RHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLSESRKYFQQLVDAVAHCH 126
Db 95 RHPHIKLYDVIKSDEIIMVIEY-AGNELFDYIVQRMKSGEQEARFFQIISAVEYCH 153

QY 127 CKGVHRDLKPENLLDNTGNLKVSDPGLSALPQEGVELLRTTCGTPNVVAPDEVLSGGY 186
Db 154 RHKIVHRDLKPENLLDNTGNLKVSDPGLSALPQEGVELLRTTCGTPNVVAPDEVLSGGY 212

QY 187 DGSAAIDWSGVLIFVLGALYLPFSETDLPGLYRKINAAEFSCPPWFAEVKFLIHRIL 246
Db 213 AGPEVDWSCGVILYVALLCGTLPFDENIPNLFKKIKGGIYTLPSHLSPGARDLIPML 272

```

QY 247 PNEKTRIOIQIKDPWRLN---YVPIRAREEYV-----LDDIRAV--- 287
 Db 273 VNPLNRISIHLMQDEWFKYDLAEVLPQDUKQEQFNKSGNEENVEIDDEWVTLK 332
 QY 288 FEGISGSVAENVERNDGPL---MMNAFEMITLSQGLNLSALFDRQD--FVKRQTRFV 342
 Db 333 TMGYDKDEIYEALESSEDTPAYNEIRNAVILIK---DNKSLIKDMQDNNVTQELDTFL 388
 QY 343 SRREPS 348
 Db 389 SOSPPET 394

RESULT 4
 ID AAK2 HUMAN STANDARD; PRT; 552 AA.
 AC P54846; Q9H1B8; Q9UD43;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 5'-AMP-activated protein kinase, catalytic alpha-2 chain (EC 2.7.1.1-)
 DE (AMPK alpha-2 chain)
 GN PRKAA2 OR AMPK.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1] SEQUENCE FROM N.A.
 RP TISSUE=Heart;
 RC MEDLINE=95047501; PubMed=7959015;
 RX Aguan K., Scott J., See C.G., Sarkar N.H.;
 RA "Characterization and chromosomal localization of the human homologue
 of a rat AMP-activated protein kinase-encoding gene: a major
 regulator of lipid metabolism in mammals.";
 RL Gene 149:345-350(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=95080410; PubMed=7988703;
 RA Beri R.K., Marley A.E., See C.G., Sopwith W.F., Aguan K., Carling D.,
 RA Scott J., Carey F.;
 RT "Molecular cloning, expression and chromosomal localisation of human
 AMP-activated protein kinase.";
 RL FEBS Lett. 356:117-121(1994).
 RN [3]
 RP SEQUENCE OF 33-552 FROM N.A.
 RA Copley V.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: RESPONSIBLE FOR THE REGULATION OF FATTY ACID SYNTHESIS
 BY PHOSPHORYLATION OF ACETYL-COA CARBOXYLASE. IT ALSO REGULATES
 CHOLESTEROL SYNTHESIS VIA PHOSPHORYLATION AND INACTIVATION OF
 HORMONE-SENSITIVE LIPASE AND HYDROXYMETHYLGLUTARYL-COA REDUCTASE.
 APPARS TO ACT AS A METABOLIC STRESS-SENSING PROTEIN KINASE
 SWITCHING OFF BIOSYNTHETIC PATHWAYS WHEN CELLULAR ATP LEVELS ARE
 DEPLETED AND WHEN 5'-AMP RISES IN RESPONSE TO FUEL LIMITATION
 AND/OR HYPOXIA. THIS IS A CATALYTIC SUBUNIT.
 CC -!- SUBUNIT: HETEROTRIMER OF A CATALYTIC SUBUNIT, A BETA AND A GAMMA
 NON-CATALYTIC SUBUNITS.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC SNF1 SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to licenses@isb-sib.ch).

 EMBL; U06454; AAA64745.1; --
 DR EMBL; AL035705; CAC17574.1; --
 DR FIR; S51025; S51025.

DR HSP; Q63450; 1A06.
 DR Genew; HGNC:9377; PRKAA2.
 DR MIM; 600497; --
 DR GO; GO:0006468; P:protein amino acid phosphorylation; TAS.
 DR GO; GO:0007165; P:signal transduction; TAS.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TKC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 KW Transferase; Serine/threonine-protein kinase; Fatty acid biosynthesis;
 KW Phosphorylation; ATP-binding; Multigene family.
 FT DOMAIN 16 268
 FT NP_BIND 22 30
 FT BINDING 45 45
 FT ACT_SITE 139 139
 FT MOD_RES 172 172
 FT CONFLICT 180 180 A -> T (IN REF. 1).
 FT CONFLICT 271 271 D -> G (IN REF. 1).
 FT CONFLICT 403 404 HL -> RQ (IN REF. 1).
 SQ SEQUENCE 552 AA; 62319 MW; C46AFCID5104975 CRC64;

Query March 29.2%; Score 670; DB 1; Length 552;
 Best Local Similarity 36.9%; Pred. No. 1.6e-40;
 Matches 151; Conservative 78; Mismatches 146; Indels 34; Gaps 7;

QY 1 MTKKMR-----RVGKYEVCRTIGETFAKVKFARNTDGDVAIKIMAKSTILKNRMVDQ 55
 Db 1 MAEQKHQGRVKIGHYVLTGVTGKVIKIGEHQLTGHRVAVILAKRQIRSLDVGK 60
 QY 56 IKREISIMKIVRHPNIVRIEVLASPSKIYIYLVETGSELPDRIVHKGRLSESRKYF 115
 Db 61 IKREIQNLKLFRRPHIILYQVISTPTDFVMVYVSGSELPDYICKHGRVEEMEARLF 120
 QY 116 QQLVDAVACHCKGVHRDLKPNLLDNGNLKVSDFGLSALPQEGVELLRTTCGTNY 175
 Db 121 QQLLSAVDYCHRRMVVHRDLKPNVLLDAMNAKTADFGLSNMDSG-EFLTSCGSPNY 179
 QY 176 VAPEVLSGGYDGSAAIDWISCGVILFVILAGVLPSETDLPGLYRKINAAEFCPPWFS 235
 Db 180 APEVLSGLYAGPEVDWISCGVILYALLCGTLPDDSHVPTLFFKIRGGVYIPEYLN 239
 QY 236 EVKFLIHRILDPNPKTRIOIQIKKDPFRINYPVIRAREE--EVNLLDTRAVFDGIE- 292
 Db 240 SVATLLMHLQVDPKRAIKDIREHEWFKQDLPSPYLPFEDPSYDANVIDDEAVKCEK 299
 QY 293 ----GSYVAENVERNDGELPMNAFEMITLSQGLNLSALFDRQDQVFKRQTRFVRRRPS 348
 Db 300 FETSEVWNLSYSGDPQQLAVAVHLI-----IDNRIMNQASEFYLAASSPS 348
 QY 349 EIANTAEAVNSWGFKSHTRNF-----KTR--LEGSSIKAGQLAV 387
 Db 349 GSPMDDSAMHIPPGLKPHRPMPLIADSPKARCPDLALNTTKPKSLAV 397

RESULT 5
 ID SNF1 CANTR STANDARD; PRT; 619 AA.
 AC O94168;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Carbon catabolite derepressing protein kinase (EC 2.7.1.1-).
 GN SNF1.
 OS Candida tropicalis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5482;
 RN [1] SEQUENCE FROM N.A.

RA Kanai T., Ogawa K., Ueda M., Tanaka A.;
 RL "Genetic evaluation of the function of SNF1 in *Candida tropicalis*."; Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: ESSENTIAL FOR RELEASE FROM GLUCOSE REPRESSION. IT INTERACTS AND HAS FUNCTIONAL RELATIONSHIP TO THE REGULATORY PROTEIN SNF4. COULD PHOSPHORYLATES CAT8 (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE NUCLEAR MEMBRANE (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -!- SNF1 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB024535; BAA75889.1; -.
 DR HSP; O63450; 1A06.
 DR InterPro; IPR000719; Prot kinase.
 DR InterPro; IPR002290; Ser Thr_kinase.
 DR Pfam; PF00069; pkinase, 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
 DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
 DR Transferase; Serine/threonine-protein kinase; ATP-binding;
 KW Phosphorylation; Carbohydrate metabolism; Nuclear protein.
 FT DOMAIN 17 29 POLY-HIS.
 FT DOMAIN 52 66 PROTEIN KINASE.
 FT NP_BIND 58 66 ATP (BY SIMILARITY).
 FT BINDING 81 81 ATP (BY SIMILARITY).
 FT ACT_SITE 174 174 BY SIMILARITY.
 FT MOD_RES 207 207 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 SQ SEQUENCE 619 AA; 70323 MW; 0FCF1FC3DCE706D7 CRC64;
 Query Match 29.0%; Score 665; DB 1; Length 619;
 Best Local Similarity 47.1%; Pred. No. 4.2e-40;
 Matches 122; Conservative 64; Mismatches 71; Indels 2; Gaps 2;
 QY 7 RVGVEVETGEGTFAKVEFARNTDGDVAIKMAKSTILKRMVMDQIKREISIMKIV 66
 DB 48 RIGRYQIITLGESEFGKVAHQVGTQCKVALKIINKLAKSDMQGRVERISILRL 107
 QY 67 RHPNIVRYELVAPSKTYIYVLEFVTGTGELFDRIVHKGRLEESERKYFOQLVDVAHCH 126
 DB 108 RHPHIIKLYDIKSDKDEIIMVTEF-AGKELFDYIVQGRKMPDEARFFQIIAAVEYCH 166
 QY 127 CKGVYHRDLKPNLLDNGMLKYSDFGLSALPGSGVELLRTCTGTNYVYAPVLSGGY 186
 DB 167 RHKIVHRDLKPNLLDQLNVKADPGLSNIMTDG-NFKTSCGSNYAAPEVISGLY 225
 QY 187 DGSAAIDVSCGVILFVILAGLYLPFSETDPLGLYRKINAAEFCPPWFSAEVKELIHRILD 246
 DB 226 AGPEVDVWSSGVILVLMGLPLPDEDEIPALFNKINSVGVTLNVLSPGAKULLTMLV 285
 QY 247 ENPKTRIQIGIKDPNFR 265
 DB 286 VNPLNRTIHEIMEDWFK 304
 RESULT 6
 SNF1_YEAST
 ID SNF1_YEAST STANDARD; PRT; 633 AA.
 AC P06782;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Carbon catabolite derepressing protein kinase (EC 2.7.1.-).
 GN SNF1 OR CAR1 OR CCR1 OR PAS14 OR GLC2 OR YDR477W OR D8035.20.

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86289463; PubMed=3526554;
 RA Celenza J.L., Carlson M.;
 RT "A yeast gene that is essential for release from glucose repression encodes a protein kinase."; Science 233:1175-1180(1986).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E., Berno A., Carpenter J., Chen E., Cherry J.M., Chung B., Duncan M., Hunnicke-Smith S., Hyman R., Kemp C., Lashkari D., Lew H., Lin D., Mossdale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X., Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N., Winant A., Yelton M., Botstein D., Davis R.W.;
 RA Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RN SEQUENCE OF 274-284; 528-539 AND 622-630, AND PHOSPHORYLATION SITE. MEDLINE=94131388; PubMed=7905477;
 RX Mitchellhill K.I., Stapleton D., Gao G., House C., Michell B., Katsis P., Witters L.A., Kemp B.E.;
 RA "Mammalian AMP-activated protein kinase shares structural and functional homology with the catalytic domain of yeast Snf1 protein kinase."; J. Biol. Chem. 269:2361-2364(1994).
 RL J. Biol. Chem. 269:2361-2364(1994).
 CC -!- FUNCTION: ESSENTIAL FOR RELEASE FROM GLUCOSE REPRESSION. IT INTERACTS AND HAS FUNCTIONAL RELATIONSHIP TO THE REGULATORY PROTEIN SNF4. INTERACTS ALSO WITH SIP1, SIP2 AND GAL83. COULD PHOSPHORYLATES CAT8.
 CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE NUCLEAR MEMBRANE.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. SNF1 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M13971; AAA35058.1; -.
 DR EMBL; U33050; AAB64904.1; -.
 DR PIR; A26030; A26030.
 DR HSP; P24941; 1HCL.
 DR SGD; S0002885; SNF1.
 DR GO; GO:0005737; C:cytoplasm; IPI.
 DR GO; GO:0005634; C:nucleus; IPI.
 DR GO; GO:0000324; C:vacuole (sensu Fungi); IPI.
 DR GO; GO:0004679; F:SNF1/AMP-activated protein kinase activity; IDA.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IDA.
 DR GO; GO:0006995; P:response to nitrogen starvation; IDA.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser Thr_kinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
 DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
 DR Transferase; Serine/threonine-protein kinase; ATP-binding;
 KW Phosphorylation; Carbohydrate metabolism; Nuclear protein.
 FT DOMAIN 18 32 POLY-HIS.
 FT DOMAIN 55 306 PROTEIN KINASE.
 FT NP_BIND 61 69 ATP (BY SIMILARITY).
 FT BINDING 84 84 ATP (BY SIMILARITY).
 FT ACT_SITE 177 177 BY SIMILARITY.
 FT MOD_RES 210 210 PHOSPHORYLATION (AUTO-).
 FT SEQUENCE 633 AA; 72045 MW; F5C63565C986C4E3 CRC64;

```
Query Match      28.9%; Score 663; DB 1; Length 633;
Best Local Similarity 39.4%; Pred. No. 6e-40;
Matches 135; Conservative 80; Mismatches 98; Indels 30; Gaps 6;

QY 8 VGVKVEVGTIGTGTGTFKAKVFAFNTDTGDNVAIKMAKSTILKNRMVDQIKREISIMKIVR 67
D 1:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
52 IGNYQIVKTLGSGSGKVLAVHTTTGQKVALKINKKVLAKSDMQGRIEREISYRLILR 111
QY 68 HPIVRLVEVLASPKIYIVLFEVYVGGELFDRIVHKGRLSESESRKYFQQLVDVAHCHC 127
D 1:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
112 HPHIILKLYDIKSDKEIIMVIEY-AGNELEFDIVQDKMSEGEARRFFQQLISAVEYCHR 170
QY 128 KGVVHRDLKPEMLLDITGNLKVSDPGI-SALPQEGVELLRITCTGPNVVAPEVLGGQYD 187
D 1:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
171 HKIVHRDLKPEMLLDEHLNVIADFGLSNIMTDG-NFLTKSCGSPNVAPEVISKLIYA 229
QY 188 GSAADINSCGVILFVILAGYLPFFSTDLPGYRKINAABFSCPPWFSAEUVKFLIHRILDP 247
D 1:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
230 GPEVDVMSGVILYVMLCRRLPFFDDESIPVLFKNISNGVYTLPKLSPGAAGLIKRMILIV 289
QY 248 NPKTRIQIGIKKDWFRNLN---YV---PIRAREEE-----VNLD 282
D 1:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
290 NPLARISHEIMQDDWFKVDLPFVYLLPDLKPHPEEENNDSDKDGSSPDNDEIDDLNV 349
QY 283 DIRAVFDGIEGSGYVAENVERNDEGPI---MNAFEMITLSQGL 322
D 1:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
350 NILSTMGVKEDEIYESLESSEDTPAFNEIRDAYMLIKENKSL 392

RESULT 7
AAK2 RAT      STANDARD;      PRT;      552 AA.
AC Q09137;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 5'-AMP-activated protein kinase, catalytic alpha-2 chain (EC 2.7.1.-)
DE (AMPK alpha-2 chain).
GN PRKA2 OR AMPK2 OR AMPK.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=94209324; PubMed=7908907;
RA Carling D., Aguan K., Woods A., Verhoeven A.J.M., Beri R.K.,
RA Brennan C.H., Sidebottom C., Davison M.D., Scott J.,
RA "Mammalian AMP-activated protein kinase is homologous to yeast and
RT plant protein kinases involved in the regulation of carbon
RT metabolism." J. Biol. Chem. 269:11442-11448 (1994).
RL J. Biol. Chem. 269:11442-11448 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Sprague-Dawley; TISSUE=Liver;
RC MEDLINE=95234757; PubMed=7719624;
RA Gao G., Widmer J., Stapleton D., Teh T., Cox T., Kemp B.E.,
RA Witters L.A.;
RT "Catalytic subunits of the porcine and rat 5'-AMP-activated protein
RT kinase are members of the SNF1 protein kinase family." J.
RL Biochem. Biophys. Acta 1266:73-82 (1995).
CC -1- FUNCTION: RESPONSIBLE FOR THE REGULATION OF FATTY ACID SYNTHESIS
CC BY PHOSPHORYLATION OF ACETYL-COA CARBOXYLASE. IT ALSO REGULATES
CC CHOLESTEROL SYNTHESIS VIA PHOSPHORYLATION AND INACTIVATION OF
CC HORMONE-SENSITIVE LIPASE AND HYDROXYMETHYLGUTARYL-COA REDUCTASE.
CC APPEARS TO ACT AS A METABOLIC STRESS-SENSING PROTEIN KINASE
CC SWITCHING OFF BIOSYNTHETIC PATHWAYS WHEN CELLULAR ATP LEVELS ARE
CC DEPLETED AND WHEN 5'-AMP RISES IN RESPONSE TO FUEL LIMITATION
CC AND/OR HYPOXIA. THIS IS A CATALYTIC SUBUNIT.
CC -1- SUBUNIT: HETEROTRIMER OF A CATALYTIC SUBUNIT, A BETA AND A GAMMA
CC NON-CATALYTIC SUBUNITS.
```

```
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=009137-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=009137-2; Sequence=VSP_004949, VSP_004950;
CC Note=Lacks the sequence parts essential for kinase activity and
CC is therefore inactive;
CC -1- TISSUE SPECIFICITY: SKELETAL MUSCLE, LOWER LEVELS IN LIVER, HEART,
CC AND KIDNEY.
CC -1- INDUCTION: BY AMP.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. SNF1
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; Z29486; CA82620.1; -;
CC EMBL; U12149; AA85033.1; -;
CC PIR; A53621; A53621.
CC HSP; Q63450; IAO61.
CC InterPro; IPR000719; Prot kinase.
CC InterPro; IPR002290; Ser Thr kinase.
CC InterPro; IPR001245; Tyr_kinase.
CC Pfam; PF00069; kinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC PRODOM; PD000001; Prot kinase; 1.
CC SMART; SM00220; S_TKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00108; PROTEIN KINASE ST; 1.
CC PROSITE; PS50011; PROTEIN KINASE DOM; 1.
CC Transfaser; Serine/threonine-protein kinase; Fatty acid biosynthesis;
CC Phosphorylation; ATP-binding; Multigene family; Alternative splicing.
CC DOMAIN 16 268 PROTEIN KINASE.
CC NP_BIND 22 30 ATP (BY SIMILARITY).
CC BINDING 45 45 ATP (BY SIMILARITY).
CC ACT_SITE 139 139 BY SIMILARITY.
CC MOD_RES 172 172 PHOSPHORYLATION (AUTO-) (POTENTIAL).
CC VARSPPLIC 32 388 Missing (in isoform Short).
CC FT VARSPPLIC 392 552 /FTid=VSP_004949.
CC FT VARSPPLIC 392 552 Missing (in isoform Short).
CC FT CONFLICT 355 355 /FTid=VSP_004950.
CC FT CONFLICT 462 462 M -> S (IN REF. 2).
CC FT CONFLICT 462 462 N -> D (IN REF. 2).
CC SQ SEQUENCE 552 AA; 62257 MW; 2829E07F674D89B1 CRC64;
Query Match      28.9%; Score 662; DB 1; Length 552;
Best Local Similarity 36.4%; Pred. No. 6e-40;
Matches 149; Conservative 79; Mismatches 147; Indels 34; Gaps 7;
QY 1 MTKMR-----RVKTEVGRTIGEGTFKAKVFAFNTDTGDNVAIKMAKSTILKNRMVDQ 55
D 1:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1 MAEKQKHGDEGVKIGHYVLGDTLGVGTGKVGKICEHQLTGKHAVKILNRQKIRSLDVVGK 60
QY 56 IKREISIMKIVRHPNIVRVLVEVLASPKIYIVLFEVYVGGELFDRIVHKGRLSESESRKYF 115
D 1:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
61 IKREIQNLKFLRPHPIIKLYQVISTPDTDFWMVEYVSGGELFDYICKGRVEEVARRLF 120
QY 116 QQLVDVAHCHCKGVYHRDLKPEMLLDITGNLKVSDPGI-SALPQEGVELLRITCTGPNY 175
D 1:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
121 QQLSADVYCHRMVVRDLKPEMLLDITGNLKVSDPGI-SALPQEGVELLRITCTGPNY 179
QY 176 VAPEVLGGCYGSAADIWSCGVILFVILAGYLPFFSTDLPGYRKINAABFSCPPWFSA 235
D 1:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
180 AAPEVISGRLYAGPEVDIWSGVILYVALLCGTLFPDDEHVPTLFFKIRGVVPIPELNR 239
QY 236 EVKFLHRLDLPNPKTRIQIGIKKDWFRNLN---YV---PIRAREEE-----VNLD 292
D 1:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

Db 240 SIATLLMHMLQVDPKRAIKIDREHEWFQKDLPSYLPBPSYDANVIDEAVKEVCEK 299

Qy 293 ----GSYVAENVENDEGFLMNAFEMITISQGLNLSALFDRQDFVKRTRFVSRRPS 348

Db 300 FECTESEVMNLSYGDPOQDLAVAYHLI-----IDNRIMNQASEFYIASSPPT 348

Qy 349 EIIANIEAVANSMGFKSHTRNF-----KTR--LEGLSSIKAGQLAV 387

Db 349 GSFWDVAMHIPPGLKHPERMPPIADSPKARCPLDALNTTKPSLAV 397

RESULT 8

ID SNF1 CANAL STANDARD; PRT; 620 AA.

AC P52497; Q00309; (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Carbon catabolite derepressing protein kinase (EC 2.7.1.-).

GN SNF1.

OS Candida albicans (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.

OX NCBI_TaxID=5476;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 32354 / B-311;

RX MEDLINE=98053924; PubMed=9393775;

RA Pether R., Chang Y.C., Kwon-Chung K.J.;

RT "A gene homologous to Saccharomyces cerevisiae SNF1 appears to be essential for the viability of Candida albicans."

RL Infect. Immun. 65:4909-4917(1997).

RN [2]

RP SEQUENCE OF 7-620 FROM N.A.

RC STRAIN=ATCC 32354 / B-311;

RA Pether R., Kwon-Chung K.J.;

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: ESSENTIAL FOR RELEASE FROM GLUCOSE REPRESSION. IT INTERACTS AND HAS FUNCTIONAL RELATIONSHIP TO THE REGULATORY PROTEIN SNF4. COULD PHOSPHORYLATE CATB (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE NUCLEAR MEMBRANE (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

CC SNF1 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; L78129; AAB48643.1; -.

DR EMBL; L39263; AAA92456.1; -.

DR HSP; Q63450; 1A06.

DR InterPro; IPR000719; Prot_Kinase.

DR InterPro; IPR002290; Ser_Thr_kinase.

DR Pfam; PF00069; pkinase; 1.

DR Prodom; PD000001; Prot_kinase; 1.

DR SMART; SM00220; S_TK; 1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.

DR PROSITE; PS00108; PROTEIN KINASE ST; 1.

DR PROSITE; PS00111; PROTEIN KINASE_DOM; 1.

DR Transferrase; Serine/threonine-protein kinase; ATP-binding;

KW Phosphorylation; Carbohydrate metabolism; Nuclear protein.

FT DOMAIN 16 29 POLY-HIS.

FT DOMAIN 53 305 PROTEIN KINASE.

FT NP_BIND 59 67 ATP (BY SIMILARITY).

FT BINDING 82 82 ATP (BY SIMILARITY).

FT ACT_SITE 175 175 BY SIMILARITY.

FT MOD_RES 208 208 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT CONFLICT 7 8 PQ -> AR (IN REF. 2).

FT CONFLICT 228 234 AGPEYDV -> SSVOKLMI (IN REF. 2).

FT CONFLICT 242 255 YVMLGRLPFDDEF -> GMSCCVVDYHSMISS (IN REF. 2).

FT CONFLICT 387 397 A -> R (IN REF. 2).

FT CONFLICT 416 416 R -> A (IN REF. 2).

FT CONFLICT 494 494 S -> L (IN REF. 2).

SQ SEQUENCE 620 AA; 70005 MW; 1806C652B5061D2B CRC64;

Query Match 28.8%; Score 661.5; DB 1; Length 620;

Best Local Similarity 44.1%; Pred. No. 7.5e-40;

Matches 128; Conservative 75; Mismatches 76; Indels 11; Gaps 6;

QY 7 RVGKYEVGRTTIGETFAKVKFARNPTDGTGNVAIKIMAKSTILKNRMVDQIKREISIMKIV 66

Db 49 RIGRYQLKTLGEGSGFKVLAHLGTCQKVALKIINKTKLAKSDMQGRVEREISYLRLL 108

QY 67 RHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGLBESERKFKYFOQLVDAVAHCH 126

Db 109 RHPIIKLYDVTKSDEIIMVIEF--AGKELFDYIVQKGMPEDEARRFQQLIAAVEYCH 167

QY 127 CKGVYHRDLKPENLLDITNGMLKVSDFGLSALPOEGVELLRTTCGTNYV--APEVLSGQG 185

Db 168 RHKIVHRDLKPENLLDITNGMLKVSDFGLSALPOEGVELLRTTCGTNYV--APEVLSGQG 226

QY 186 YDGSRAADIWSCGVILFVILAGVLPSETDLPGLYKINAAEFSCPPWFSAEVKFLIHRIL 245

Db 227 YAGPEVDVMSAGVILYVMLGRLPFDDEFIPALFKKISNGVYTLFNYLSAGAKHLITRL 286

QY 246 DNPXTRIQIQIKKDPWFRLL---NVV---PIRAREEEVNLD--TRAV 287

Db 287 VYNPLNRTIHEIMEDDFKQDMPDYLPPDLSSKNKSKIDVDEVDVIRAL 336

RESULT 9

AAK1 HUMAN STANDARD; PRT; 550 AA.

ID AAK1 HUMAN STANDARD; PRT; 550 AA.

AC Q13131; Q00286; Q9UNQ4;

DT 15-JUL-1998 (Rel. 36, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE 5'-AMP-activated protein kinase, catalytic alpha-1 chain (EC 2.7.1.-) (AMPK alpha-1 chain).

DE PRKAA1 OR AMPK1.

GN Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Mammary gland;

RA Yano K.;

RT "Nucleotide sequence of cDNA for human AMP-activated protein kinase alpha-1.";

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Blood;

RX MEDLINE=20499367; PubMed=11042152;

RA Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G., Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W., Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.;

RT "Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells.";

RL Genome Res. 10:1546-1560(2000).

RN [3]

RP SEQUENCE OF 27-200 FROM N.A.

RC TISSUE=Intestine;

RA Taboada E.N., Hickey D.A.;

RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE OF 294-550 FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=96132781; PubMed=8557660;
 RA Stapleton D., Michellhill K.I., Gao G., Widmer J., Michell B.J.,
 RA Ten T., House C.M., Fernandez C.S., Cox I., Witters L.A.,
 RA Kemp B.E.;
 RT "Mammalian AMP-activated protein kinase subfamily.";
 RL J. Biol. Chem. 271:611-614(1996).
 CC !- FUNCTION: RESPONSIBLE FOR THE REGULATION OF FATTY ACID SYNTHESIS
 CC BY PHOSPHORYLATION OF ACETYL-COA CARBOXYLASE. IT ALSO REGULATES
 CC CHOLESTEROL SYNTHESIS VIA PHOSPHORYLATION AND INACTIVATION OF
 CC HORMONE-SENSITIVE LIPASE AND HYDROXYMETHYLGLUTARYL-COA REDUCTASE.
 CC APPEARS TO ACT AS A METABOLIC STRESS-SENSING PROTEIN KINASE
 CC SWITCHING OFF BIOSYNTHETIC PATHWAYS WHEN CELLULAR ATP LEVELS ARE
 CC DEPLETED AND WHEN 5'-AMP RISES IN RESPONSE TO FUEL LIMITATION
 CC AND/OR HYPOXIA. THIS IS A CATALYTIC SUBUNIT (BY SIMILARITY).
 CC !- SUBUNIT: HETEROTRIMER OF A CATALYTIC SUBUNIT, A BETA AND A GAMMA
 CC NON-CATALYTIC SUBUNITS
 CC !- PTM: AUTOPHOSPHORYLATED (BY SIMILARITY).
 CC !- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC SNF1 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB022017; BAA36547.1; -;
 CC EMBL; AF100763; AAD43027.1; -;
 CC EMBL; U22456; AAG64850.1; -;
 CC EMBL; Y12856; CAA73361.1; -;
 CC F1R; G01743; G01743.
 CC HSSP; Q63450; 1A06.
 CC Genew; HGNC:9376; PRKAA1.
 CC MIM; 602739; -;
 CC DR GO; GO:0005622; C:intracellular; IC.
 CC DR GO; GO:0008189; P:apoptosis inhibitor activity; NAS.
 CC DR GO; GO:0004691; P:AMP-dependent protein kinase activity; NAS.
 CC DR GO; GO:0042557; P:eukaryotic elongation factor-2 kinase activ. .; ISS.
 CC DR GO; GO:0000187; P:activation of MAPK; NAS.
 CC DR GO; GO:0046318; P:negative regulation of glucosylceramide bio. .; NAS.
 CC DR GO; GO:0017148; P:negative regulation of protein biosynthesis; ISS.
 CC DR GO; GO:0045768; P:positive regulation of anti-apoptosis; NAS.
 CC DR GO; GO:0045542; P:positive regulation of cholesterol biosynth. .; NAS.
 CC DR GO; GO:0046321; P:positive regulation of fatty acid oxidation; ISS.
 CC DR GO; GO:0045722; P:positive regulation of gluconeogenesis; ISS.
 CC DR GO; GO:0046326; P:positive regulation of glucose import; ISS.
 CC DR GO; GO:0006468; P:protein amino acid phosphorylation; ISS.
 CC DR GO; GO:0001666; P:response to hypoxia; NAS.
 CC DR InterPro; IPR000719; Prot_kinase.
 CC DR InterPro; IPR002290; Ser_thr_kinase.
 CC DR Pfam; PF00069; pkinase; 1.
 CC DR PRINTS; PR00109; TYRKINASE.
 CC DR ProDom; PD000001; Prot_kinase; 1.
 CC DR SMART; SM00220; S_TK; 1.
 CC DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 CC DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
 CC DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
 CC KW Transferase; Serine/threonine-protein kinase; Fatty acid biosynthesis;
 CC Phosphorylation; ATP-binding.
 CC FT DOMAIN 18 270 PROTEIN KINASE.
 CC FT NP_BIND 24 32 ATP (BY SIMILARITY).
 CC FT BINDING 47 47 ATP (BY SIMILARITY).
 CC FT ACT_SITE 141 141 BY SIMILARITY.
 CC FT MOD_RES 174 174 PHOSPHORYLATION (AUTO-) (POTENTIAL).
 CC FT CONFLICT 28 28 T -> A (IN REF. 3).
 CC FT CONFLICT 193 193 A -> V (IN REF. 3).
 CC FT CONFLICT 199 199 I -> L (IN REF. 3).
 CC FT CONFLICT 260 260 S -> T (IN REF. 2).
 CC SEQUENCE 550 AA; 62793 MW; 3316183D744BE325 CRC64;

Query Match 28.6%; Score 655.5; DB 1; Length 550;
 Best Local Similarity 32.0%; Pred. No. 1.7e-39;
 Matches 148; Conservative 97; Mismatches 159; Indels 59; Gaps 6;

 QY 7 RVGKYEVGRITGEGTFAKVFARNTDTGDNVAIKMAKSTILKRMVDQIKREISIMKIV 66
 DB 14 KIGHYILGDTLGVGTGKVGKGBELTGHKVAVKILNRQKIRSLDVVGKIRREIQNLKF 73
 QY 67 RHPNIVRLVEVLASPSKIYIVLFTVGGELPDRIVHKGRLESRSKYFOOLVDAVAHCH 126
 DB 74 RHPHIKIYQVISTPESDIFWMVEYVSGGELFDYICNKRGLDEKSRRLFOQILSGVDYCH 133
 QY 127 CRGVVHRDLKPEINLLDINGNLKVDGFLSALPOEGVELLRTTCTGPNYVAPEVLSGGCY 186
 DB 134 RHMVVRDLKPEINLLDAHNAKIADFLGSLNMSDG-EFLRTSCGSPNYAAPEVLSGLY 192
 QY 187 DGSADIWCSGVILFVLVLAGYLPSETDLPGLYKINAAEFSCPPWFSAEYKFLHRLD 246
 DB 193 AGPEVDIWSGGVILYALLCGTLFPDDHVTLPFKICDGIFFTPQYLNPSVISLKHMLQ 252
 QY 247 PNPKRIRIQIGIKDPWFRLNVPIRAREE-----EENVLDIDIRAVFDGIEGYSVAE 298
 DB 253 VDPMKRASKIDREHEWFKQDLPKLYFPEDPSYSTMIDDEALKVECFECSEEEVLSC 312
 QY 299 NVRENDEGLMNAPEMTLSQGL-----NLSALFDRQDFVKR 337
 DB 313 LYNNRHQDPLAV-AVYHLIDNRRIINNEAKDFYATSPDPSFLDDHHLTPRHPERVPLVA 371
 QY 338 QT-----RFVSRERSEIIANTEAVANSNGFSKSTRN- 369
 DB 372 ETPRARHTLDELNPQSKHQGVKAKWHLGTRSQSRPNIDMAEVCRAIKQLDYEWKVNVP 431
 QY 370 FKTRLEGLSSIKAGLAVVIEIYVAPSLFVMDVVRKAAGETLE 412
 DB 432 YYLVRERKNPVTSTYSKMSQLYQVDSRTYLLDFRSIDDEITE 474

 RESULT 10
 ID AAKI_RAT STANDARD; PRT; 548 AA.
 AC P54645;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 5'-AMP-activated protein kinase, catalytic alpha-1 chain (EC 2.7.1.-)
 DE (AMPK alpha-1 chain).
 GN PRKAA1 OR AMPK1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]_TaxID=10116;
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=Sprague-Dawley; TISSUE=Hypothalamus, and Liver;
 RX MEDLINE=96132781; PubMed=8557660;
 RA Stapleton D., Michellhill K.I., Gao G., Widmer J., Michell B.J.,
 RA Ten T., House C.M., Fernandez C.S., Cox T., Witters L.A.,
 RA Kemp B.E.;
 RT "Mammalian AMP-activated protein kinase subfamily.";
 RL J. Biol. Chem. 271:611-614(1996).
 RN [2]
 RP PARTIAL SEQUENCE.
 RC STRAIN=Sprague-Dawley; TISSUE=Hypothalamus, and Liver;
 RX MEDLINE=95050763; PubMed=7961907;
 RA Stapleton D., Gao G., Michell B.J., Widmer J., Michellhill K.,
 RA Ten T., House C.M., Witters L.A., Kemp B.E.;
 RT "Mammalian 5'-AMP-activated protein kinase non-catalytic subunits are
 RT homologs of proteins that interact with yeast Snf1 protein kinase.";
 RL J. Biol. Chem. 269:29343-29346(1994).
 CC !- FUNCTION: RESPONSIBLE FOR THE REGULATION OF FATTY ACID SYNTHESIS
 CC BY PHOSPHORYLATION OF ACETYL-COA CARBOXYLASE. IT ALSO REGULATES
 CC CHOLESTEROL SYNTHESIS VIA PHOSPHORYLATION AND INACTIVATION OF
 CC HORMONE-SENSITIVE LIPASE AND HYDROXYMETHYLGLUTARYL-COA REDUCTASE.

RESULT 13

QY 127 CKGVYHRLDKPENLLDTNGNLKVSDFGLSALFQEGVELLRTCTGTPNYVAPEVLGGQY 186

Db 140 NHMIVHRLDKPENLLDANKNIKTIADFGLSNYMTDG-DLLSTACGSPNYAAPLISNKLY 198

QY 167 DGSRAADTSCGVILFVILAGVLPSEFDTLGLYRKINAEEFSCPPWFSABVKFLIHRILD 246

Db 199 VGPEVDLWSCGVILYAMLCGLTLPDDQNVTLTFAKISGRYTVFYSMEKQAADLSTMLQ 258

QY 247 PNPKTRIQIGIKKDPWFLN- ---YVPIRAREEEVNLDDIRAV 287

Db 259 VDEKRVADVKRIVNHSFRIDLPPYLPFECENESSIVDIDWQSV 303

RESULT 14

KKKL_YEAST

ID KKKL_YEAST STANDARD; PRT; 1518 AA.

AC P34244;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Probable serine/threonine-protein kinase YKL101W (EC 2.7.1.-).

GN YKL101W OR YKL453.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=94078677; PubMed=8256524;

RA Pallier C., Valens M., Puzos V., Fukuhara H., Cheret G., Sor F.,

RA Bolotin-Fukuhara M.;

RT "DNA sequence analysis of a 17 kb fragment of yeast chromosome XI

RT physically localizes the MRB1 gene and reveals eight new open reading

RT frames, including a homologue of the KIN1/KIN2 and SNF1 protein

RT kinases.";

RL Yeast 9:1149-1155(1993).

CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

CC NIM1 SUBFAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC EMBL; X71133; CAA50456.1; -.

DR EMBL; Z28101; CAA81941.1; -.

DR PIR; S37928; S37928.

DR HSSP; Q63450; 1A06.

DR SGD; S0001584; HSL1.

DR GO; GO:0005935; C:bud neck; IDA.

DR GO; GO:0005940; C:septin ring; IDA.

DR GO; GO:0000086; P:G2/M transition of mitotic cell cycle; IGI.

DR GO; GO:0006468; P:protein amino acid phosphorylation; IDA.

DR GO; GO:0000074; P:regulation of cell cycle; IMP.

DR GO; GO:0000135; P:septin checkpoint; IGI.

DR InterPro; IPR000719; Prot_kinase.

DR InterPro; IPR002290; Ser_Thr_kinase.

DR InterPro; IPR001245; Tyr_kinase.

DR Pfam; PF00069; pkinase; 1.

DR PRINTS; PR0109; TYRKINASE.

DR ProDom; PD000001; Prot_kinase; 1.

DR SMART; SM00220; S_TKc; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.

DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.

KW Hypothetical protein; transferase; Serine/threonine-protein kinase;

FT ATP-binding. 81 369 PROTEIN KINASE.

FT DOMAIN

QY 7 RVGKVEGVGTIGEGTFAKVPARNTDTGDAVKIMAKSTILKRMVMDQIKREISIMKIV 66

Db 20 KLGNFVKEITCKGFACVAKRTHQTGYDAIKILNCRMKGLGTWNKTRNEIDNLQKL 79

QY 67 RHPTNIVRLVNLASPSKIIYVLEFVTGGLFDRIHVHKGRLSEESRKYFQOLDAVAHCH 126

Db 80 THEPHITRLFRVSTSPSDIPLNVLVSGGLFSYITRKGLALPIRSRRYFOOILSGVSCH 139

QY 127 CKGVYHRLDKPENLLDTNGNLKVSDFGLSALFQEGVELLRTCTGTPNYVAPEVLGGQY 186

Db 140 NHMIVHRLDKPENLLDANKNIKTIADFGLSNYMTDG-DLLSTACGSPNYAAPLISNKLY 198

QY 167 DGSRAADTSCGVILFVILAGVLPSEFDTLGLYRKINAEEFSCPPWFSABVKFLIHRILD 246

Db 199 VGPEVDLWSCGVILYAMLCGLTLPDDQNVTLTFAKISGRYTVFYSMEKQAADLSTMLQ 258

QY 247 PNPKTRIQIGIKKDPWFLN- ---YVPIRAREEEVNLDDIRAV 287

Db 259 VDEKRVADVKRIVNHSFRIDLPPYLPFECENESSIVDIDWQSV 303

RESULT 14

KKKL_YEAST

ID KKKL_YEAST STANDARD; PRT; 1518 AA.

AC P34244;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Probable serine/threonine-protein kinase YKL101W (EC 2.7.1.-).

GN YKL101W OR YKL453.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=94078677; PubMed=8256524;

RA Pallier C., Valens M., Puzos V., Fukuhara H., Cheret G., Sor F.,

RA Bolotin-Fukuhara M.;

RT "DNA sequence analysis of a 17 kb fragment of yeast chromosome XI

RT physically localizes the MRB1 gene and reveals eight new open reading

RT frames, including a homologue of the KIN1/KIN2 and SNF1 protein

RT kinases.";

RL Yeast 9:1149-1155(1993).

CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

CC NIM1 SUBFAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC EMBL; X71133; CAA50456.1; -.

DR EMBL; Z28101; CAA81941.1; -.

DR PIR; S37928; S37928.

DR HSSP; Q63450; 1A06.

DR SGD; S0001584; HSL1.

DR GO; GO:0005935; C:bud neck; IDA.

DR GO; GO:0005940; C:septin ring; IDA.

DR GO; GO:0000086; P:G2/M transition of mitotic cell cycle; IGI.

DR GO; GO:0006468; P:protein amino acid phosphorylation; IDA.

DR GO; GO:0000074; P:regulation of cell cycle; IMP.

DR GO; GO:0000135; P:septin checkpoint; IGI.

DR InterPro; IPR000719; Prot_kinase.

DR InterPro; IPR002290; Ser_Thr_kinase.

DR InterPro; IPR001245; Tyr_kinase.

DR Pfam; PF00069; pkinase; 1.

DR PRINTS; PR0109; TYRKINASE.

DR ProDom; PD000001; Prot_kinase; 1.

DR SMART; SM00220; S_TKc; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.

DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.

KW Hypothetical protein; transferase; Serine/threonine-protein kinase;

FT ATP-binding. 81 369 PROTEIN KINASE.

FT DOMAIN

QY 7 RVGKVEGVGTIGEGTFAKVPARNTDTGDAVKIMAKSTILKRMVMDQIKREISIMKIV 66

Db 20 KLGNFVKEITCKGFACVAKRTHQTGYDAIKILNCRMKGLGTWNKTRNEIDNLQKL 79

QY 67 RHPTNIVRLVNLASPSKIIYVLEFVTGGLFDRIHVHKGRLSEESRKYFQOLDAVAHCH 126

Db 80 THEPHITRLFRVSTSPSDIPLNVLVSGGLFSYITRKGLALPIRSRRYFOOILSGVSCH 139

QY 127 CKGVYHRLDKPENLLDTNGNLKVSDFGLSALFQEGVELLRTCTGTPNYVAPEVLGGQY 186

Db 140 NHMIVHRLDKPENLLDANKNIKTIADFGLSNYMTDG-DLLSTACGSPNYAAPLISNKLY 198

QY 167 DGSRAADTSCGVILFVILAGVLPSEFDTLGLYRKINAEEFSCPPWFSABVKFLIHRILD 246

Db 199 VGPEVDLWSCGVILYAMLCGLTLPDDQNVTLTFAKISGRYTVFYSMEKQAADLSTMLQ 258

QY 247 PNPKTRIQIGIKKDPWFLN- ---YVPIRAREEEVNLDDIRAV 287

Db 259 VDEKRVADVKRIVNHSFRIDLPPYLPFECENESSIVDIDWQSV 303

RESULT 14

KKKL_YEAST

ID KKKL_YEAST STANDARD; PRT; 1518 AA.

AC P34244;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Probable serine/threonine-protein kinase YKL101W (EC 2.7.1.-).

GN YKL101W OR YKL453.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=94078677; PubMed=8256524;

RA Pallier C., Valens M., Puzos V., Fukuhara H., Cheret G., Sor F.,

RA Bolotin-Fukuhara M.;

RT "DNA sequence analysis of a

```

FT NP_BIND      87 95 ATP (BY SIMILARITY).
FT BINDING     110 110 ATP (BY SIMILARITY).
FT ACT_SITE    239 239 BY SIMILARITY.
SQ SEQUENCE    1518 AA; 169592 MW; 803F84F751241DD CRC64;

Query Match      25.7%; Score 589.5; DB 1; Length 1518;
Best Local Similarity 30.7%; Pred. No. 3.2e-34;
Matches 138; Conservative 90; Mismatches 135; Indels 87; Gaps 10;

QY 2 TKKWR-----VGKYEVRTTGGTFAKVPARNTDGDVNAVKIMAKSTIL----- 48
DQ 67 TKSKRRSRDVTGPKLGTGLKSGSGRVRKAKNMTGQLAAIKIVPKKAFVHCSSNGT 126
QY 49 -----KRRWDQ-----IKREISMKIVRPENIVRLVYEVLASP 81
DQ 127 VPNSYSSSMVTSNVSSPSIASREHNSQTPYGIEREIVIMKLSHTNVMALEFVWENK 186
QY 82 SKIYVLFVFTGGELFDRIVHKGLESESRKYQQLVDAVAHCHCKGVHRDLKPENLL 141
DQ 187 SELVLEVDGGLFDVLSVSKGLPEREAHYFKQIVGVSCHSFNI CHRDLPENLL 246
QY 142 LD-TGNLKVSDPGLSALPQGVBLRTTCTGPNYVAPVLSGGQYDGSAAIWSGCVIL 200
DQ 247 LDKNRRIKIADFMAALEFN-KLLKTSQSPHYASPEIWMGRPYHGSPDWSGCVIL 305
QY 201 FVILAGLPFSETDPLGLYRKINAAPFCPPWESAERKFLIHRITLDNPKTRIOIQIKK 260
DQ 306 FALLTGHLPNDNDNIKLLKLVQSGKYOMPNSLSSEARDLISKLVLDPEKRTTQELK 365
QY 261 DPMFR-----LNVYPIRABE-----BEVNLDDI 284
DQ 366 HPLIKKYDPLPVNKLVRKMRKNMARGKNSDLHLNNVSPSIVTLHSGKEIDSLRSL 425
QY 285 RAVFDGEGYVAENVERN--DEGLPMNAFEMITLSQGLNLSALFDRDFVKRQRFV 342
DQ 426 QILWHGVSRRLITAKLQKPMSEKFLYSLLOVKORHSISLSSSEN-----KKSATES 480
QY 343 SREPSPIIANIEAVANSMGFKSHTRNFKT 372
DQ 481 SVNEPR--IYASKTANTGLRSENNDVKT 508

RESULT 15
MRX4_HUMAN
ID MRX4_HUMAN STANDARD; PRT; 752 AA.
AC Q96L34; Q96JG7; Q9BYD8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE MAP/microtubule affinity-regulating kinase 4 (EC 2.7.1.27)
DE (MAP/microtubule affinity-regulating kinase like 1).
GN MARK4 OR MARKL1 OR KIAA1860.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RA Drewes G., Mandelkow E.M.;
RT "MARK4, homologue of MARK1, MARK2 and MARK3.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RA Kato T., Satoh S., Okabe H., Kitahara O., Ono K., Kihara C.,
RA Tanaka T., Tsunoda T., Yamaoka Y., Nakamura Y., Furukawa Y.;
RT "Isolation of a novel human gene, MARKL1, homologous to MARK3 and its
RL involvement in hepatocellular carcinogenesis.";
RN Neoplasia 3:4-9(2001).
RN [3]
SEQUENCE FROM N.A. (ISOFORM 1).

```

```

RC TISSUE=Brain;
RA Beghini A., Magnani I., Roversi G., Piepoli T., DiTerlizzi S.,
RA Pollo B., Conti A.M.F., Cowell J.K., Finocchiaro G., Iarizza L.;
RT "Neural progenitor-restricted isoform of MARKL1 gene is upregulated by
RT 19q13 amplification in human glioblastoma.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RA MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 8:85-95(2001).
RN [5]
SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Thyroid;
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project.";
RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -|- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q96L34-1; Sequence=Displayed;
CC Name=2; Synonyms=MARKL1S;
CC IsoId=Q96L34-2; Sequence=VSP_004946;
CC -|- TISSUE SPECIFICITY: Ubiquitous. Isoform 2 is brain-specific.
CC -|- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. MARK
CC SUBFAMILY.
CC -|- SIMILARITY: Contains 1 kinase-associated (KAL) domain.
CC -|- SIMILARITY: Contains 1 UBA domain.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
EMBL; AY057448; AAL23683.1; -
EMBL; AB049127; BAB39380.1; -
EMBL; AY120867; AM55491.1; -
EMBL; AK073272; BAC11510.1; -
EMBL; AB058763; BAB47489.1; ALT_INIT.
HSSP; Q63450; 1A06.
Genew; HGNC:13538; MARK4.
MIM; 606495; -
InterPro; IPR001772; Kinase Cterm.
InterPro; IPR000719; Prot_kinase.
InterPro; IPR002290; Ser_thr_kinase.
InterPro; IPR001245; Tyr_kinase.
InterPro; IPR000449; UBA_domain.
Pfam; PF02149; KAL; 1.
Pfam; PF00069; Pkinase; 1.
Pfam; PF00827; UBA; 1.
PRINTS; PR00109; TYRKINASE.
ProDom; PD000001; Prot_kinase; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
PROSITE; PS50030; UBA; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding;
KW Alternative splicing.
FT DOMAIN 59 310 PROTEIN KINASE.
FT DOMAIN 324 368 UBA.
FT DOMAIN 703 752 KAL.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 16:48:42 ; Search time 19 seconds
(without alignments)
2257.433 Million cell updates/sec

Title: US-09-824-735-2

Perfect score: 2293

Sequence: 1 MTKMRRVGKVEVGRITIGEG.....IINRATEGIPKSBILRTITF 446

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1484.5	64.7	445	T09903	serine/threonine-s
2	1381.5	60.2	480	A86427	probable serine/th
3	1333	58.1	440	T14735	probable serine/th
4	1314	57.3	441	C84667	probable protein k
5	1313	57.3	440	T14736	probable serine/th
6	1234.5	53.8	453	G86141	protein T25K16.13
7	1040	45.4	441	E85362	hypothetical prote
8	1009.5	44.0	445	T50802	serine/threonine p
9	1006.5	43.9	456	T48478	serine/threonine p
10	999.5	43.6	421	E96522	hypothetical prote
11	983.5	42.9	489	T04862	probable serine/th
12	979.5	42.7	421	T48202	protein kinase AK2
13	978	42.7	520	G86414	probable protein k
14	968.5	42.2	461	T14822	probable serine/th
15	925	40.8	502	T02306	probable protein k
16	924	40.3	469	E84644	probable protein k
17	900.5	39.3	439	T03444	protein kinase hom
18	879	38.3	435	E84707	probable protein k
19	869.5	37.9	426	C71408	probable protein k
20	869	37.9	442	T48203	hypothetical prote
21	845	36.9	527	A53467	protein kinase SNF
22	808.5	35.3	431	T02496	probable protein k
23	718	31.3	472	E90100	SNF-related kinase
24	705	30.7	511	A56009	serine/threonine-s
25	701.5	30.6	576	T41587	probable carbon c
26	700.5	30.5	512	JC1446	serine/threonine-s
27	686	29.9	504	T10449	probable serine/th
28	677	29.5	512	T52633	serine/threonine-s
29	676	29.5	504	T07415	probable serine/th

ALIGNMENTS

RESULT 1

T09903

serine/threonine-specific protein kinase homolog T22A6.230 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C;Accession: T09903
R;Bevan, M.; Zimmermann, W.; Gruenisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, June 1999

A;Reference number: Z16896

A;Accession: T09903

A;Molecule type: DNA

A;Residues: 1-445 <BEV>

A;Cross-references: EMBL:AL078637; GSPDB:GN00062; ATSP:T22A6.230

A;Experimental source: cultivar Columbia; BAC clone T22A6

C;Genetics:

A;Gene: ATSP.T22A6.230

A;Map position: 4

A;Intons: 53/3; 74/3; 98/3; 134/3; 159/3; 177/3; 219/3; 249/3; 290/3; 329/3; 368/3; 387

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolc

Query Match 64.7%; Score 1484.5; DB 2; Length 445;

Best Local Similarity 62.3%; Pred. No. 2.4e-56;

Matches 273; Conservative 84; Mismatches 80; Indels 1; Gaps 1;

QY	5	MRVGVGVGRIGSTGTAQKVPARTGTGDNVAIKIMAKSTILKRVAVDQIKREISIMK	64
DB	3	VRKVGKVELGRTIGSTGTAQKVPARTGTGDNVAIKIMAKSTILKRVAVDQIKREISIMK	62
QY	65	IVRHPNIVRLYEVLASPSKIYIVLEFVITGGELFDRIVHKGRLSESESKYFQQLVDVAH	124
DB	63	LVRRPCVRLYEVLASPSKIYIVLEFVITGGELFDRIVHKGRLSESESKYFQQLVDVAH	122
QY	125	CHCKGVYHDLKPENLLDTNGNLKVSDFGLSALPQEGVELLRTCTGPNVAVDEVLSGQ	184
DB	123	CHSKGVYHDLKPENLLDTNGNLKVSDFGLSALPQEGVELLRTCTGPNVAVDEVLSGQ	182
QY	185	GYDGSAAIDVWSGVILFVILAGYLPFSETDPLGLVRKINAAEFSCPPWFSAEVFLHRI	244
DB	183	GYNGVADVWSGVILFVILAGYLPFSETDPLGLVRKINAAEFSCPPWFSAEVFLHRI	242
QY	245	LDPNPKTRITQIGIKKDPWFRNLNYPFIRARBEVEEYVNLDDIRAVFDIGSGVVAENVERND	304
DB	243	LDPNPKTRITQIGIKKDPWFRNLNYPFIRARBEVEEYVNLDDIRAVFDIGSGVVAENVERND	301
QY	305	EGFLMNAEMITLSQGLNLSALFDRQDFVKKQTRFVSRREPSEIINIANIVANSNGFK	364
DB	302	TGPLTINAEFDLILSQGLNLSALFDRQDFVKKQTRFVSRREPSEIINIANIVANSNGFK	361
QY	365	SHTRNFKTLEGISGTAKQALAVWIEIYEVAPSLFMDVRKAAGETLTYHYFKYKLSKL	424
DB	362	THIRNYKMEVEGLSANKTSFHSVILFVAVPSILMDVIONAGDAEYLAIFYKTFCSKL	421

QY 425 ENIWRATEGIPKSEILR 442
::||: : :: :
Db 422 DDIWKPPDASMRNRVTK 439

RESULT 2

A86427
probable serine/threonine kinase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: A86427
R;Authors: Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; PMID:21016719; PMID:11130712
A;Accession: A86427
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-480 <STO>
A;Cross-references: GB:AE005172; NID:gl1120780; PIDN:AAG30961.1; GSPDB:GNC0141
C;Genetics:
A;Map position: 1
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

RESULT 3

T14735
probable serine/threonine kinase (EC 2.7.1.-) SNFL1 - sorghum
N;Alternate names: SNF1 protein kinase homolog SNFL1
C;Species: Sorghum bicolor (sorghum)

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
 C;Accession: T14735
 R;Annen, F.; Stockhaus, J.
 Plant Mol. Biol. 36, 529-539, 1998
 A;Title: Characterization of a Sorghum bicolor gene family encoding putative protein kin
 A;Reference number: Z18177; MUID:98145442; PMID:9484448
 A;Accession: T14735
 A;Status: preliminary; translated from GB/EMBL/DDBL
 A;Molecule type: DNA
 A;Residues: 1-440 <ANN>
 A;Cross-references: EMBL:Y12454; NID:G2632251; PIDN:CAA73067.1; PID:G2632252
 A;Experimental source: cultivar TX 430
 C;Genetics:
 A;Gene: SNFL1
 C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
 C;Keywords: phosphotransferase; serine/threonine-specific protein kinase

RESULT 4

C84657
probable protein kinase [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C/Accession: C84667
M./Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A./Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A./Reference number: A84420; MUID:20083487; PMID:10617197
A./Accession: C84667
A./Status: preliminary
A./Molecule type: DNA
A./Residues: 1-441 <STO>
A./Cross-references: GB:AE002093; NID:G3885328; PIDN:AAC77856.1; GSPDB:GN00139
C./Genetics:
A./Gene: At2g26980

A;Map position: 2

A; Map position: 2
C; Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

Query Match 57.3%; Score 1314; DB 2; Length 441;
Best Local Similarity 57.7%; Pred. NO. 3.8e-49;
Matches 248; Conservative 81; Mismatches 95; Indels 6; Gaps 3;

QV 6 RRVGKYEVGRTIGEGTFAKVKFARNTDGDNVAIKIMAKSTILKNRMVDOIKREISIMKI 65

姓名	性别	年龄	籍贯	民族	文化程度	职业	婚姻	子女	健康状况	宗教信仰	政治面貌	特长	其他
王德胜	男	45	山东	汉族	高中	教师	已婚	2	良好	无	中共党员	书法	
李小红	女	38	河南	汉族	初中	工人	已婚	1	良好	无	群众	舞蹈	
张小明	男	25	江苏	汉族	大学	程序员	未婚	0	良好	无	共青团员	编程	
赵国强	男	52	四川	汉族	小学	农民	已婚	3	一般	无	群众	农活	
陈丽娟	女	41	广东	汉族	高中	护士	已婚	2	良好	无	群众	护理	
刘伟强	男	33	浙江	汉族	大学	工程师	已婚	1	良好	无	中共党员	设计	
孙小芳	女	28	湖北	汉族	初中	售货员	未婚	0	良好	无	群众	销售	
周大伟	男	48	湖南	汉族	高中	司机	已婚	2	良好	无	群众	驾驶	
吴美玲	女	35	福建	汉族	大学	医生	已婚	1	良好	无	群众	医疗	
郑国强	男	55	安徽	汉族	小学	工人	已婚	3	一般	无	群众	体力	
林小华	女	22	江西	汉族	高中	学生	未婚	0	良好	无	共青团员	学习	
黄大刚	男	42	广西	汉族	初中	农民	已婚	2	良好	无	群众	农活	
徐丽娟	女	37	山西	汉族	高中	教师	已婚	1	良好	无	群众	教学	
周伟强	男	30	陕西	汉族	大学	程序员	未婚	0	良好	无	共青团员	编程	
吴小芳	女	26	云南	汉族	初中	售货员	未婚	0	良好	无	群众	销售	
郑大伟	男	46	贵州	汉族	高中	司机	已婚	2	良好	无	群众	驾驶	
林美玲	女	34	海南	汉族	大学	医生	已婚	1	良好	无	群众	医疗	
周国强	男	54	重庆	汉族	小学	工人	已婚	3	一般	无	群众	体力	
吴小华	女	23	四川	汉族	高中	学生	未婚	0	良好	无	共青团员	学习	
黄大刚	男	43	湖南	汉族	初中	农民	已婚	2	良好	无	群众	农活	
徐丽娟	女	36	湖北	汉族	高中	教师	已婚	1	良好	无	群众	教学	
周伟强	男	29	浙江	汉族	大学	程序员	未婚	0	良好	无	共青团员	编程	
吴小芳	女	27	广东	汉族	初中	售货员	未婚	0	良好	无	群众	销售	
郑大伟	男	47	河南	汉族	高中	司机	已婚	2	良好	无	群众	驾驶	
林美玲	女	33	山东	汉族	大学	医生	已婚	1	良好	无	群众	医疗	
周国强	男	53	江苏	汉族	小学	工人	已婚	3	一般	无	群众	体力	
吴小华	女	24	江西	汉族	高中	学生	未婚	0	良好	无	共青团员	学习	
黄大刚	男	44	广西	汉族	初中	农民	已婚	2	良好	无	群众	农活	
徐丽娟	女	35	山西	汉族	高中	教师	已婚	1	良好	无	群众	教学	
周伟强	男	31	陕西	汉族	大学	程序员	未婚	0	良好	无	共青团员	编程	
吴小芳	女	28	云南	汉族	初中	售货员	未婚	0	良好	无	群众	销售	
郑大伟	男	48	贵州	汉族	高中	司机	已婚	2	良好	无	群众	驾驶	
林美玲	女	34	海南	汉族	大学	医生	已婚	1	良好	无	群众	医疗	
周国强	男	54	重庆	汉族	小学	工人	已婚	3	一般	无	群众	体力	
吴小华	女	23	四川	汉族	高中	学生	未婚	0	良好	无	共青团员	学习	
黄大刚	男	43	湖南	汉族	初中	农民	已婚	2	良好	无	群众	农活	
徐丽娟	女	36	湖北	汉族	高中	教师	已婚	1	良好	无	群众	教学	
周伟强	男	29	浙江	汉族									

DB 9 RRVGKYEVGRITIGETFAKVVFARNSETGEPVALKILDKEVLKHQMAEQIRREIATMKL 68

QY 66 VRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESESRYFOOLVDAVAHC 125

Question	Answer
1. The first step in the process of creating a new product is to identify a market need.	True
2. A product that is new to the market and new to the company is called a new product.	True
3. A product that is new to the market but not new to the company is called a modified product.	True
4. A product that is not new to the market but is new to the company is called a new product.	True
5. A product that is not new to the market and not new to the company is called an existing product.	True
6. The process of creating a new product is called product development.	True
7. The process of creating a new product is called product innovation.	True
8. The process of creating a new product is called product design.	True
9. The process of creating a new product is called product engineering.	True
10. The process of creating a new product is called product manufacturing.	True
11. The process of creating a new product is called product distribution.	True
12. The process of creating a new product is called product promotion.	True
13. The process of creating a new product is called product pricing.	True
14. The process of creating a new product is called product placement.	True
15. The process of creating a new product is called product packaging.	True
16. The process of creating a new product is called product labeling.	True
17. The process of creating a new product is called product advertising.	True
18. The process of creating a new product is called product sales.	True
19. The process of creating a new product is called product support.	True
20. The process of creating a new product is called product feedback.	True

Db 69 IKHPNVQLYEVMSKTKIFILEYVTGGELFDKI VNDGRMKEDEARRYFQQLIHAVDYC 128

QY 126 HCKGVYHRDLKPENLLDTNGNLKVSDFGLSALPOEGVE--LLRTTCGTPNYVAPEVLSG 183

[illegible]

Db 129 HSRGVYHRDLKPENLLDSYGNLKISDFGLSALSQVRDDGLLHTSCGTPNYVAPEVLND 188

QY 184 QGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHR 243

Country	Year	Value
Algeria	1980	1.00
Algeria	1981	1.00
Algeria	1982	1.00
Algeria	1983	1.00
Algeria	1984	1.00
Algeria	1985	1.00
Algeria	1986	1.00
Algeria	1987	1.00
Algeria	1988	1.00
Algeria	1989	1.00
Algeria	1990	1.00
Algeria	1991	1.00
Algeria	1992	1.00
Algeria	1993	1.00
Algeria	1994	1.00
Algeria	1995	1.00
Algeria	1996	1.00
Algeria	1997	1.00
Algeria	1998	1.00
Algeria	1999	1.00
Algeria	2000	1.00
Algeria	2001	1.00
Algeria	2002	1.00
Algeria	2003	1.00
Algeria	2004	1.00
Algeria	2005	1.00
Algeria	2006	1.00
Algeria	2007	1.00
Algeria	2008	1.00
Algeria	2009	1.00
Algeria	2010	1.00
Algeria	2011	1.00
Algeria	2012	1.00
Algeria	2013	1.00
Algeria	2014	1.00
Algeria	2015	1.00
Algeria	2016	1.00
Algeria	2017	1.00
Algeria	2018	1.00
Algeria	2019	1.00
Algeria	2020	1.00
Algeria	2021	1.00
Algeria	2022	1.00
Algeria	2023	1.00
Algeria	2024	1.00
Algeria	2025	1.00
Algeria	2026	1.00
Algeria	2027	1.00
Algeria	2028	1.00
Algeria	2029	1.00
Algeria	2030	1.00
Algeria	2031	1.00
Algeria	2032	1.00
Algeria	2033	1.00
Algeria	2034	1.00
Algeria	2035	1.00
Algeria	2036	1.00
Algeria	2037	1.00
Algeria	2038	1.00
Algeria	2039	1.00
Algeria	2040	1.00
Algeria	2041	1.00
Algeria	2042	1.00
Algeria	2043	1.00
Algeria	2044	1.00
Algeria	2045	1.00
Algeria	2046	1.00
Algeria	2047	1.00
Algeria	2048	1.00
Algeria	2049	1.00
Algeria	2050	1.00
Algeria	2051	1.00
Algeria	2052	1.00
Algeria	2053	1.00
Algeria	2054	1.00
Algeria	2055	1.00
Algeria	2056	1.00
Algeria	2057	1.00
Algeria	2058	1.00
Algeria	2059	1.00
Algeria	2060	1.00
Algeria	2061	1.00
Algeria	2062	1.00
Algeria	2063	1.00
Algeria	2064	1.00
Algeria	2065	1.00
Algeria	2066	1.00
Algeria	2067	1.00
Algeria	2068	1.00
Algeria	2069	1.00
Algeria	2070	1.00
Algeria	2071	1.00
Algeria	2072	1.00
Algeria	2073	1.00
Algeria	2074	1.00
Algeria	2075	1.00
Algeria	2076	1.00
Algeria	2077	1.00
Algeria	2078	1.00
Algeria	2079	1.00
Algeria	2080	1.00
Algeria	2081	1.00
Algeria	2082	1.00
Algeria	2083	1.00
Algeria	2084	1.00
Algeria	2085	1.00
Algeria	2086	1.00
Algeria	2087	1.00
Algeria	2088	1.00
Algeria	2089	1.00
Algeria	2090	1.00
Algeria	2091	1.00
Algeria	2092	

DB 189 RGYDGA TADMWSCGVLYVLLAGYLPFD DSNLMNLYKKISSGFE NCPFWLSLGAMK LITR 248

Qy 244 ILDPNPKTRIQIOGIKKDPWFRNLNYVPIRAREEEVNLDIDRAVFDGIEGSYVAENVN 303

[illegible]

249 ILDPNPMTRVTPQEVFEDEWFKKDYKPPVFEERDDSNMDDIDAVFKDSEHLV---TEKR 305

QY 304 DEGPLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANSMGF 363

[illegible]

DB 306 EEQPAAINAFEIISMSRGLNLENLFDPEQEF-KRETRITLRGGANEIIEKIEEAAKPLGF 364

QY 364 KSHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMVDVRKAAGETLEYHKFYKLCCK 423

[illegible]

DB 365 DVQKKNYKMRLENVKAGRKGNLNVATETIQVAPSLAHMVQVSKSKGDTLEFFHKFYKLSNS 424

Qy 424 LENIIWRATE 433

100

DB 425 LEQVVWINNE 434

RESULT 5
E1422C

TI4/36
probable serine/threonine kinase (EC 2.7.1. -) SNET.2 - sorghum

N;Alternate names: SNF1 protein kinase homolog SNFL2

C;Species: Sorghum bicolor (sorghum)

C:\Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:\Accession: T14736

R;Annen, F.; Stockhaus, J.

Plant Mol. Biol. 36, 529-539, 1998

A/Title: Characterization of a Sorghum bicolor gene family encoding putative
A:Reference number: Z18177. MIMD:98145442. PMID:9484448

A;Accession: T14736

A;Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA
A:Residues: 1-440 <ANN>

A;Cross-references: EMBL:Y12465; NID:g2632253; PIDN:CAA73068.1; PID:g2632254

A;Experimental source: cultivar TX 430

C:Genetics:
A:Gene: SNFL2

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinases

C;Keywords: phosphotransferase; serine/threonine-specific protein kinase

Query Match 57.3%: Score 1313: DB 2: Length 440:

QY 243 RILDNPXTRIOGIQIKDPHRLNLYVPIR-AREEEVNLLDIRAVDPGIEGYSVAENVE 301
 DB 255 RILENPITRISIALLEDEWFKYKPPSPDQDEDDITDDVDAFNSKECLVTEKKE 314
 QY 302 RNDGGLPMNAFEMITTSQGLNLSALFDRRQDFVKRQTRFVSRPSEIITANIEAVANS 361
 DB 315 K---PVSMAFELISSSEFSLNLFKQAQLVKKTRFTSQRSASEINMSKMEETAKPL 370
 QY 362 GFKSHTRNFKRLRGLSSIKAGOLAVVIEYVAPSLFMDVVRKAAGETLEYHK---FY 417
 DB 371 GFNVKONIKYKIMKGDGSRGKQLSVATEFVAPSLHVVLRKTDGDTLEFHKVCDSFY 430
 QY 418 KKLCSKLENIIV 429
 DB 431 KNFSSGLKDVVW 442

RESULT 7
 B85362
 Hypothetical protein AT4G30960 [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cross)
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001
 C;Accession: B85362
 R;Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
 Nature 402, 769-777, 1999
 A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
 A;Reference number: A85001; MUID:20083488; PMID:10617198
 A;Accession: B85362
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-441 <STO>
 A;Cross-references: GB:NC_001268; MID:g7269998; PIDN:CAB79814.1; GSPDB:GN00140
 A;Gene: AT4G30960
 A;Map position: 4
 C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

Query Match 45.4%; Score 1040; DB 2; Length 441;
 Best Local Similarity 49.0%; Pred. No. 1.4e-37;
 Matches 217; Conservative 73; Mismatches 109; Indels 44; Gaps 10;

QY 9 GKYEVRITGEGTFAKVFARNTDGTGNVAIKMAKSTILKNRMVDDQIKREISIMKIVRH 68
 DB 22 GEYELGRLLGHCTFAKVYHARNIQTKSVAMKVGVKGVGVNDQIKREISVMRVKH 81
 QY 69 PNIVELYVLASPKIYIVLBFVTGGEILFDRIVHKGRLESESRKYFOQLVDVAHCHCK 128
 DB 82 PNIVELHEVMASKSIYFAMELVREGELFAKVA-KGRLEDVARVYFQQLISAVDFCHSR 140
 QY 129 GYVHRDLKPENLILLDTNGLKVSDFGLSA---LPOEGVELLRTTCGTPNYVAPEVLSGQ 184
 DB 141 GYVHRDLKPENLILLDEGNLKVTDPLGSAFTEHLKQDG--LLHTTCGTPAYVAPEVILKK 198
 QY 185 GYDGSADINSGVILFVLGAYLPFSETDPLGYRKINAABFSCPPWFSAEVKFLHRI 244
 DB 199 GYDGAKADLWSGVILFVLGAYLPFQDNLVNMYRKIYRGDFKCPGWLSSDARRLVTKL 258
 QY 245 LDPNPKTRIOGIQIKDPWF-----RLNYPVIRA---REEEVLDDIRAVDFDIEGSYV 296
 DB 259 LDPNPNTRITTEKVMDSFWFKQAQTRSNPEVPAATITTEEDVDP----- 303
 QY 297 AENVNDEGGLPMNAFEMITTSQGLNLSALFDRRQDFVKRQTRFVSRPSEIITANIEA 356
 DB 304 --LVHKSKEETETLNAFHIALSEGFLDPLFEEKKKEKREMRPATSRPASSVISLEE 361
 QY 357 VANSWGFKSHTR-NFKTRLEGSLIKAGOLAVVIEYVAPSLFMDVVRKAAGETLEYH 414
 DB 362 AAR-VGNKFDVRSRSRVRIEKGQRKGKLAVEAEIIFAVAPSFVWVEVKDHDGDTLEYN 420
 QY 415 KFYKKLCS-----KLENIIVWAT 432
 DB 421 NF-----CSTALRPALXDFWTST 439

RESULT 8

T50802

serine/threonine protein kinase-like protein - Arabidopsis thaliana

N;Alternate names: protein T30N20_200

C;Species: Arabidopsis thaliana (mouse-ear cross)

C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 18-Aug-2000

C;Accession: T50802

R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew

submitted to the Protein Sequence Database, July 2000

A;Reference number: Z25240

A;Accession: T50802

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-445 <BEV>

A;Cross-references: EMBL:AL365234

A;Experimental source: cultivar Columbia; BAC clone T30N20

C;Genetics:

A;Map position: 5

A;Note: T30N20_200

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

Query Match 44.0%; Score 1009.5; DB 2; Length 445;

Best Local Similarity 47.8%; Pred. No. 2.8e-36;

Matches 213; Conservative 81; Mismatches 111; Indels 41; Gaps 11;

QY 6 REV--GKYEVRITGEGTFAKVFARNTDGTGNVAIKMAKSTILKN-RMVDQIKREISI 62

DB 5 RRVLFKYEGRLLGKGTFAKVYKKEIIGGECVAIKVINKDQVMKRPQMWEIKREISI 64

QY 63 MKIVSHPNIVRLYVLAFLSPSKIYIVLFEVTGGEILFDRIVHKGRLESESRKYFOQLVDV 122

DB 65 MKLVHRPNIVLKEVYNATKTIFFVMEFVKGGEILFCKI-SKGKLHEDAAIRYFOQLISAV 123

QY 123 AHCCHKGVVHRDLKPENLILLDTNGLKVSDFGLSALPOEGVZ--LLRTTCGTPNYVAPEV 180

DB 124 DYCHSGVSHRDLKPENLILLDEGNLKLISDFGLSALPEQLDGLLHTTCGTPAYVAPEV 183

QY 181 LSGQVDSGAADINSGVILFVLGAYLPFSETDPLGYRKINAABFSCPPWFSAEVKFL 240

DB 184 LKKGYDGAKADINSGVILFVLGAYLPFQDNLVNMYRKIIFRADPEFFWFSPEARLL 243

QY 241 THRILDPNPKTRIOGIQIKDPWFNLNYPVIRA-----REEEVLDDIR 286

DB 244 ISKLLVDDPDRISIPALMRTWLNRKNTPLAFKIDPFCISQSSKNNEEEDGD---- 299

QY 287 VFDGIEGYSVAENVNDEGGLPMNAFEMI-TLSQGLNLSALFDRRQDFVKRQTRFVSR 345

DB 300 -----CEN-QTEPISPKFFNAFFIFISMSSGFLDLSLFESKR---KVQSVFTGRS 345

QY 346 EPSEIITANIEAVANSNGFK-SHTRNFKTRLEGSLIKAGOLAVVIEYVAPSLFMDVVR 404

DB 346 STATEWKELETVTKEMNMKVKTQDFKVKMEGKTEGRKGRLSMTAEVEVAPEISVVEFC 405

QY 405 KAAGETLEYHKFY-KKLCSKLENIIV 429

DB 406 KSAGDTLEYDLYEEVEVRPALNDIVW 431

RESULT 9

T48478

serine/threonine protein kinase-like protein - Arabidopsis thaliana

N;Alternate names: protein T28J14.10

C;Species: Arabidopsis thaliana (mouse-ear cross)

C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000

C;Accession: T48478

R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lee

submitted to the Protein Sequence Database, April 2000

A;Reference number: Z24493

A;Accession: T48478

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-456 <BEV>

Db 200 KGYDAKVDIWSGCVILFVLMAGVLPFDNRNVMYKKIYRGPFCRPFSTELTRLLSK 259

QY 244 ILDPNPKTRIQIQIKDPPRLNVPVR-----AREEEVNLDIRAVFGIEG 293

Db 260 LLETNPKRFTFPBIMENSWMFKGFKHIFKVEDDKLVVDDLESDSVESDRDASAAS 319

QY 294 SYVAENVE-RNDEG---PLMNAPEMITLSQGLNLSALFDRQDFVKRQTRFVSRREPS 348

Db 320 ESEILEYPRRVRGGLPRPASLNAPDIISFGQFDLSGLFDDGQ---GSRFVSGAPVS 375

QY 349 EIANIRAVANSMGFKSHTRNFKTRLGLSSIKAGQLAVIIEYVAPSLFMDVVRKAAG 408

Db 376 KIISKLEBIAKWSFTVRKDCRVSLGSRQGVKGLPTIAAEIPELTPSLVWVEVKKGG 435

QY 409 ETLEYHKF-YKLCSCLENI 427

Db 436 DKTEYDFCNELPKLQNL 455

RESULT 12

T48202

protein kinase AK22 (BC 2.7.1.1-) - Arabidopsis thaliana

N:Alternate names: protein T20L15.80

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 17-Nov-2000

C/Accession: T48202; S66335; S58261

R:By: Van, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24488

A:Accession: T48202

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-421 <BEV>

A:Cross-references: EMBL:AL162351

A:Experimental source: cultivar Columbia; BAC clone T20L15

R:Thiemler, F.; Kirchner, M.; Teuber, R.; Ditttrich, P.

Plant Mol. Biol. 29, 551-565, 1995

A:Title: Differential accumulation of the transcripts of 22 novel protein kinase genes

A:Reference number: S66314; MUID:96123233; PMID:8534852

A:Accession: S66335

A:Molecule type: DNA

A:Cross-references: EMBL:X86967; NID:g928911; PID:CAA60530.1; PID:g928912

C:Genetics:

A:Gene: AK22

A:Map position: 5

A:Note: T20L15.80

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

C:Keywords: ATP; phosphotransferase; protein kinase

F:136-193/Domain: protein kinase homology (fragment) <KIN>

Query Match 42.7%; Score 979.5; DB 2; Length 421;

Best Local Similarity 49.8%; Pred. No. 4.9e-35;

Matches 206; Conservative 71; Mismatches 116; Indels 21; Gaps 7;

QY 10 KYEVGRTTGGTFAKVKPARNTDGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVRHP 69

Db 11 RYEVGKFLGQGTFAKVVHARHLKTGDSVAIKVIDKERLLKVGTEQIKRHSARLLRHP 70

QY 70 NIVRLYEVLASPKTIYVLEFVTGSELFDRIHVKGRLEESRSKYFQOLVDVAHCHCKG 129

Db 71 NIVELHEVMATSKIIYFVMEHVKGELFNK-VSTGKLREDAVKRYFOQLVRAVDFCHSRG 129

QY 130 VYHRDLKPENLLDNGNLKYSDFGLSALP---QEGVELLRTTCGTPNVVAPVLSGQ 185

Db 130 VCHRLKPENLLDDEHGNLKSDFGLSALSRSRQDG--LHTTCGTPAVVAPVLSRNG 187

QY 186 YDGSAAIDWSCGVILFVILAGYLPFSETDLPGLYRKINAABFSCPPWFSAEVKFLIHRIL 245

Db 188 YDGFKADWVSCGVILFVILAGYLPFRDNLNMELYKIGKAEVKFPNWLAPGAKRLKRL 247

QY 246 DPNPKTRIQIQIKDPPRLNVPVRAREEEVNLDIRAVFGIEGSAENVERDE 305

Db 248 DPNPKTRIVSTKIMKSSWFRKG-----LQEEVKESVEEETVDAEAGNASAEKKEK-- 300

QY 306 GLPMNNAPEMITLSQGLNLSALFDRQDFVKRQTRFVSRREPSIIIANIEAVANSMGPKS 365

Db 301 -CINLNAPEIISLSTGFDLSGLFEKGE--KEEMRFTSNREASEITEKLEIVIGDKOLMKY 357

QY 366 HVRNFKTRLEGSLSSIKAGQLAVIIEYVAPSLFMDVVRKAAGTLEVHKYKK 419

Db 358 RKKEHEWVK---MSAEATVVEAEVFIAPSYHNVVLKSGGDTAEYKRWYKE 407

RESULT 13

G86414

probable protein kinase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002

C/Accession: G86414

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G86414

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-520 <STO>

A:Cross-references: GB:AE005172; NID:g9502417; PID:AAF8116.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

Query Match 42.7%; Score 978; DB 2; Length 520;

Best Local Similarity 47.0%; Pred. No. 6.8e-35;

Matches 208; Conservative 82; Mismatches 121; Indels 32; Gaps 10;

QY 8 VKGYEVGRTTGGTFAKVKPARNTDGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVR 67

Db 71 MKYELGKLLGHGTFAKVLQNTKSGDKVAIKVIDKERIMKSGLVAKIKREISILRVR 130

QY 68 HPIVRLYEVLASPKTIYVLEFVTGSELFDRIHVKGRLEESRSKYFQOLVDVAHCHC 127

Db 131 HPYIVLHEVMATSKIIYFVMEYVGGELFNTVA-KGRLPETATRRYQQIISVSFCHG 189

QY 128 KGVYHRDLKPENLLDNGNLKVSDFGLSA---LPQSGVELLRTTCGTPNVVAPVLSG 183

Db 190 RGVYHRDLKPENLLDNGNLKVSDFGLSANAQLRQDG--LCHTFCGTPAVIAPVLR 247

QY 184 QGYDGAADWSCGVILFVILAGYLPFSETDLPGLYRKINAABFSCPPWFSAEVKFLIHR 243

Db 248 KGYDAADWVSCGVILFVILAGYLPFYDKNIMVMYKIKYGEFCRPNWFSDDLRLTR 307

QY 244 ILDPNPKTRIQIQIKDPPRLNVPVR-----AREEEVNLDIRAVFGIEGSA 295

Db 308 LLDTPDPTRITTPETIMKNRWFKKGFVKFYIEDKLCREDED---ESEEASSG-RSST 363

QY 296 VAE-----NVERNDEG---PLMNAPEMITLSQGLNLSALFDRQDFVKRQTRFVSRRE 346

Db 364 VESDAEPDVKMGIGMSPRPSLNAPDIISFGSGFDLSGLFEKGE--EGTRFVSGAP 420

QY 347 PSEITANTIEAVANSMGFKSHTRNFKTRLGLSSIKAGQLAVIIEYVAPSLFMDVVRKA 406

Db 421 VSKIIISKLEBIAKIVSFVTRKKEWLSRLEGCRGAKGLPTIAAEIFELTPSLVWVEVKKK 480

QY 407 AGETLEYHKF-YKLCSCLENI 428

Db 481 GGRREYEFCKELRPLEKLI 503

RESULT 15
T02306
probable protein kinase [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein F13P17.2
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C;Accession: T02306; D84753
R;Roundley, S.D.; Linn, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
A;Submitted to the EMBL Data Library, July 1998
A;Description: Arabidopsis thaliana chromosome II BAC F13P17 genomic sequence.
A;Reference number: Z14657
A;Accession: T02306

A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-502 <R0>
A;Cross-references: EMBL:AC004481; NID:g3337347; PIDN:AAC27394.1; PID:g3337349
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: D84753
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-502 <STO>
A;Cross-references: GB:AE002093; NID:g3337349; PIDN:AAC27394.1; GSPDB:GN00139
C;Genetics: At2g34180; F13P17.2
A;Gene: At2g34180; F13P17.2
A;Map position: 2
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
P;55-311/Domain: protein kinase homology <KIN>

Query Match 40.8%; Score 935; DB 2; Length 502;
Best Local Similarity 44.7%; Pred. No. 4.3e-33;
Matches 199; Conservative 91; Mismatches 123; Indels 32; Gaps 11;

QY 10 KYEVGTIGEGTFAKVKFARNITDTGNVAIKIMAKSTILKNRMVDOIKREISIMKIVRHP 69
DB 56 KYEIGKLLGHGSPAKVYLARNITHSGEDVAIKVIDKEKIVKSGLAGHIKREISILRRVRHP 115
QY 70 NIVRLYEVLASPKIYIVLEFVTGGELFDRIVHKGLEESSEKYPQQLVDVAHCHCKG 129
DB 116 YIVHLLVAVATKIYIVMEYVREGELYNTVA-RGRLEGTARYFQQLISSVAFCHSRG 174
QY 130 VYHRDIKPENLLDNTGNLKVSDFGLSA---LPQGVELLRTTCGTPNVAVAEVLSGQG 195
DB 175 VYHRDUKLENLLDDNGNVKVSDFGVSUVEQJQEGI--CQTFCTGTPAYLADEVITRKG 232
QY 186 YDGSAAADISGCGVILEVILAGYLPFSSETDLPGLYRKINAAEFSCPPWFSAEKNFLIHRIL 245
DB 233 YEGAKADISGCGVILFVLMAGYLPFDKNIILVMYTKIYKQFKCPKWFSPELARLVTRML 292
QY 246 DPNPKTRIQGHKKDPWFLNVVPIR-----AREEEVNLDIDRAVDFDIEGSYVA 297
DB 293 DTPNEDTRITPELKHWRVFKKGFHKVRYENDKLCREDDNDNDSSSSLSG-RSSTAS 351
QY 298 E-----NVRNDE--GELMMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFSRRBFSFI 350
DB 352 EGDAAEDFDIKVDGMPRPASLNAFDILSFS---DLSLGFIEEG----GQGARFVSAAPMTKI 404
QY 351 IANIEAVANSWGFKSHTRNPKTEPLEGLSSIKACQAVVLIIEYVAPSLFMDVYRKAAGET 410
DB 405 ISKLEETAKEKVENWKKDMSVRECEGAKGFLIRVIFELTFSLVVVEVKKGGNI 464
QY 411 LEYHKF-YKCLSKLENII-WRATE 433
DB 465 EEEYEFCKNELRPOLEKLMHYQADE 489

Search completed: October 20, 2003, 16:52:20
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 16:52:02 ; Search time 69 seconds
(without alignments)

1059.522 Million cell updates/sec

Title: US-09-824-735-2

Perfect score: 2293

Sequence: 1 MTKKRRVGVKVEGRTIGEG.....IIWRAEGIPKSEILRTIP 446

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 609560 seqs, 16391702 residues

Total number of hits satisfying chosen parameters: 609560

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/2/pubpa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2293	100.0	446	9	US-09-824-735-2
2	1545	67.4	450	12	US-10-292-408-9
3	669	29.2	552	9	US-09-824-735-4
4	663	28.9	633	9	US-09-824-735-3
5	663	28.9	633	10	US-09-801-368-338
6	608.5	26.5	1203	9	US-09-799-875-5
7	602	26.3	778	12	US-10-354-358-92
8	602	26.3	778	12	US-10-116-326-2
9	593.5	25.9	651	9	US-09-870-937-10
10	593.5	25.9	651	10	US-09-974-298-112
11	593.5	25.9	651	12	US-10-354-358-8
12	591	25.8	793	15	US-10-195-101-32
13	589.5	25.7	1518	10	US-09-801-368-152
14	588	25.6	795	10	US-09-919-585-12
15	588	25.6	795	12	US-10-161-565-25

16	588	25.6	795	15	US-10-142-356-9	Sequence 9, Appli
17	586.5	25.6	798	11	US-09-823-187-95	Sequence 95, Appl
18	586.5	25.6	832	10	US-09-919-585-21	Sequence 21, Appl
19	586	25.6	752	10	US-09-835-081-2	Sequence 2, Appli
20	585	25.5	744	10	US-09-919-585-3	Sequence 3, Appli
21	584	25.5	729	12	US-10-161-565-26	Sequence 26, Appl
22	584	25.5	729	15	US-10-142-356-11	Sequence 11, Appl
23	584	25.5	729	15	US-10-195-101-33	Sequence 33, Appl
24	583	25.4	744	10	US-09-835-081-4	Sequence 4, Appli
25	579.5	25.3	779	8	US-08-817-832B-31	Sequence 31, Appl
26	578	25.2	688	12	US-10-161-565-28	Sequence 28, Appl
27	578	25.2	688	12	US-10-161-565-29	Sequence 29, Appl
28	578	25.2	713	12	US-10-161-565-27	Sequence 27, Appl
29	576.5	25.1	776	11	US-09-823-187-92	Sequence 92, Appl
30	576.5	25.1	776	11	US-09-823-187-93	Sequence 93, Appl
31	576.5	25.1	776	15	US-10-195-101-34	Sequence 34, Appl
32	575	25.1	691	10	US-09-919-585-6	Sequence 6, Appli
33	575	25.1	722	8	US-08-817-832B-32	Sequence 32, Appl
34	575	25.1	724	10	US-09-919-585-9	Sequence 9, Appli
35	575	25.1	745	12	US-10-161-565-24	Sequence 24, Appl
36	575	25.1	745	15	US-10-195-101-36	Sequence 36, Appl
37	574	25.0	668	14	US-10-054-579-2	Sequence 2, Appli
38	574	25.0	668	15	US-10-195-072-2	Sequence 2, Appli
39	574	25.0	668	15	US-10-195-071-2	Sequence 2, Appli
40	574	25.0	674	15	US-10-283-247-2	Sequence 2, Appli
41	573	25.0	783	9	US-09-815-915-2	Sequence 2, Appli
42	573	25.0	783	12	US-10-393-316-2	Sequence 2, Appli
43	570	24.9	674	10	US-09-842-582-2	Sequence 2, Appli
44	570	24.9	674	15	US-10-283-247-7	Sequence 7, Appli
45	570	24.9	674	15	US-10-283-247-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1

US-09-824-735-2
; Sequence 2, Application US/09824735
; Patent No. US20020095032A1
; GENERAL INFORMATION:
; APPLICANT: ZHU, JIAN-KANG
; APPLICANT: LIU, JIPING
; APPLICANT: ISHITANI, MANABU
; APPLICANT: HALFTER, URSULA
; APPLICANT: KIM, CHEOL-SOO
; TITLE OF INVENTION: PROTEINS AND DNA RELATED TO SALT TOLERANCE IN PLANTS
; FILE REFERENCE: 205645US20
; CURRENT APPLICATION NUMBER: US/09/824,735
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 60/824,735
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-824-735-2

Query Match 100.0%; Score 2293; DB 9; Length 446;
Best Local Similarity 100.0%; Pred. No. 2.4e-193;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MTKKRRVGVKVEGRTIGEGTFAKVKFARNDTGDNVAIKIMAKSTILKNRVDQIKREI	60
DB	1	MTKKRRVGVKVEGRTIGEGTFAKVKFARNDTGDNVAIKIMAKSTILKNRVDQIKREI	60
QY	61	SIIMKIVRPNVRLVYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKVFQQLVD	120
DB	61	SIIMKIVRPNVRLVYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKVFQQLVD	120
QY	121	AVAHCHCKGVHDKPENLALDTGNLKVSDGILALPQGVGELLRTTCCTPNVAPEV	180

Db 121 AVAHCHCKGVYHRDLKPENLLDITNGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEV 180
QY 181 LSGGCGYDGSADINWSCGVILPFIAGYLPFSETDLPGLYRKINAAEFCPPWFSAEYKFL 240
Db 181 LSGGCGYDGSADINWSCGVILPFIAGYLPFSETDLPGLYRKINAAEFCPPWFSAEYKFL 240
QY 241 IHRILDPNPKTRIQIGIKKDPWFLNVPTRAREEVEVNLDITRAVFDGIEGSSVAENV 300
Db 241 IHRILDPNPKTRIQIGIKKDPWFLNVPTRAREEVEVNLDITRAVFDGIEGSSVAENV 300
QY 301 ERNDEGLPMNAFEMITLSQGLNLSALFDRRQDFVKQTRFVSRREPSEIIANIEAVANS 360
Db 301 ERNDEGLPMNAFEMITLSQGLNLSALFDRRQDFVKQTRFVSRREPSEIIANIEAVANS 360
QY 361 MGKSHTRNFKTRLEGLSSIKAGQLAVIIEYVAPSLFMDVVRKAAGETLVEYHKFYKFL 420
Db 361 MGKSHTRNFKTRLEGLSSIKAGQLAVIIEYVAPSLFMDVVRKAAGETLVEYHKFYKFL 420
QY 421 CSKLENIINWATEGIPKSEILRTITF 446
Db 421 CSKLENIINWATEGIPKSEILRTITF 446

RESULT 2
US-10-292-408-9
; Sequence 9, Application US/10292408
; Publication No. US20030182692A1
; GENERAL INFORMATION:
; APPLICANT: VAN THIELEN, NOCHA
; APPLICANT: CHEN, RUOYING
; TITLE OF INVENTION: PROTEIN KINASE STRESS-RELATED POLYPEPTIDES AND METHODS
; FILE REFERENCE: 16313-0178
; CURRENT APPLICATION NUMBER: US/10/292,408
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: 60/346,096
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Physcomitrella patens
US-10-292-408-9

Query Match 67.4%; Score 1545; DB 12; Length 450;
Best Local Similarity 66.0%; Pred. No. 1.5e-127;
Matches 293; Conservative 70; Mismatches 75; Indels 6; Gaps 4;
QY 5 MRVGVYEVGRTIGEGTFAKVKFARNITDGDNVAIKIMAKSTILKNRMVDQIKREISIMK 64
Db 3 MRVGVYEVGRTIGEGTFAKVKFARNITDGDNVAIKIMAKSTILKNRMVDQIKREISIMK 62
QY 65 IVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHGRLEESBSRKYFOOLVDAVAH 124
Db 63 LVRHPNIVRLHEVLASRCKIYIILEFVTGGELFDKIVHGRLEENDSRKYFOOLMDGVY 122
QY 125 CHCKGVYHRDLKPENLLDITNGNLKVSDFGLSALPQEGVE--LLRTTCGTPNYVAPEVLS 182
Db 123 CHSGVSHRDLKPENLLDITNGNLKVSDFGLSALPQEVREDGLLHTTCGTPNYVAPEVLN 182
QY 183 GQGYDGSADINWSCGVILPFIAGYLPFSETDLPGLYRKINAAEFCPPWFSAEYKFLH 242
Db 183 DKGYDAVADINWSCGVILPFIAGYLPFSETDLPGLYRKINAAEFCPPWFSAEYKFLH 242
QY 243 RILDPNPKTRIQIGIKKDPWFLNVPTRAREEVEVNLDITRAVFDGIEGSSVAENV 302
Db 243 NILDPNPKTRIRMRGIRDDEWFKNVVPMYDDEINLDVETAFDDSKQFVKEQREV 302
QY 303 NDEGLPMNAFEMITLSQGLNLSALFDRRQDFVKQTRFVSRREPSEIIANIEAVANS 362
Db 303 KDVGPSLNAFELISLQGLNLSALFDRRQDFVKQTRFVSRREPSEIIANIEAVANS 362

QY 363 FKSHTRNFKTRLEGLSSIKAGO-LAVVIEIYVAPSLFMDVVRKAAGETLVEYHKFYKFLC 421
Db 363 FGVTGRNYKVRLEAASECRISQHLAVALEYVEVAPSLFMDVVRKAAGETLVEYHKFYKFLC 422
QY 422 SKLENIINWATEGIPKSEILRTIT 445
Db 423 TRLKDIW--TTAVDKDEV-KTLT 443

RESULT 3
US-09-824-735-4
; Sequence 4, Application US/09824735
; Patent No. US20020095032A1
; GENERAL INFORMATION:
; APPLICANT: ZHU, JIAN-KANG
; APPLICANT: LIU, JIPIING
; APPLICANT: ISHITANI, MANABU
; APPLICANT: HALFTER, URSULA
; APPLICANT: KIM, CHEOL-SOO
; TITLE OF INVENTION: PROTEINS AND DNA RELATED TO SALT TOLERANCE IN PLANTS
; FILE REFERENCE: 205645US20
; CURRENT APPLICATION NUMBER: US/09/824,735
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 60/824,735
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-824-735-4

Query Match 29.2%; Score 669; DB 9; Length 552;
Best Local Similarity 36.9%; Pred. No. 2.4e-50;
Matches 151; Conservative 77; Mismatches 147; Indels 34; Gaps 7;
QY 1 MTKQMR-----RVGKYEVGRTIGEGTFAKVKFARNITDGDNVAIKIMAKSTILKNRMVDQ 55
Db 1 MAEQKHQDGRVKIGHYVLGDTLGVGTGFKVKGIGEHQLTGHKVAVKILNQRSLDVGK 60
QY 56 IKREISIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHGRLEESBSRKYF 115
Db 61 IKREIQNLKLFRRPHIILKYQVISTPTDFVMVMEYVSGGELEFDYICKHGRVEEMEARL 120
QY 116 QQLVDAVAHCHCKGVYHRDLKPENLLDITNGNLKVSDFGLSALPQEGVELLRTTCGTPNY 175
Db 121 QQLSADVYCHRMVVRDLKPENLLDITNGNLKVSDFGLSALPQEGVELLRTTCGTPNY 179
QY 176 VAPEVLSGGYDGSADINWSCGVILPFIAGYLPFSETDLPGLYRKINAAEFCPPWFSA 235
Db 180 AAPEVTSRGLYAGPEVDIWSGCVILVALLCGTLPFDDHVPFLFKIRGVPYIPEYLN 239
QY 236 EYKFLIHRILDPNPKTRIQIGIKKDPWFLNVPTRAREE--EVNLDITRAVFDGIE- 292
Db 240 SVATLEMLMLQVDPDLKRTATIKDIREHWFKQGLPFLPDPSPYDANIDDEAVKEVCEK 299
QY 293 ----GSYVAENVNDEGLPMNAFEMITLSQGLNLSALFDRRQDFVKQTRFVSRREP 348
Db 300 FECTESEVNSLYSGDPQQLAVAYHLI-----IDNRRIMQASEFYLASSPPS 348
QY 349 EIANIEAVANSMGFKSHTRNF-----KTR--LEGISITKAGQLAV 387
Db 349 GSFMDDSAMHIPPGLKPHPERMPPLIADSPKARCPDLALNTTKPSLAV 397

RESULT 4
US-09-824-735-3
; Sequence 3, Application US/09824735
; Patent No. US20020095032A1
; GENERAL INFORMATION:
; APPLICANT: ZHU, JIAN-KANG


```

QY 127 CKGYVHRDLKPNLLDTNGNLKYSDFGLSALPOBQVHLLRTTCGTNYVAPEVLSGGY 186
Db 123 CRNVHRDLKAEKLNLLDANLKIADFGSNLFTFG-QLLKTWCGSPPYAAPLFEQKEY 181
QY 187 DGSAAIDWSCGVILFVILAGVLPFSETDPLCLYRKINAABSCPPWFSAEVKFLIHRILD 246
Db 182 DGPKVDIWSLGVLYLVLCGALPDGSLTQNLARVLISGKRIIPFFMSTCEHLIRHMLV 241
QY 247 PNKTRIQIQIKKDPWFL-----NVVPIRA-----REEEV---NLDDIRAVFD-GIE 292
Db 242 LDPNKLRLSMEQICKHKKWMLGDADPNFDRLIABCOQLKEERQVDPDLNEDVLLAMEDGLD 301
QY 293 GSYVAENVERNDEBPLM-----MNAFEMITLSQGLNLSALFPRDRQFVKRQRFV 342
Db 302 KE---QTLQABQAGTAMNISVPQVLINPQNIIVEPDPGLNLDS--DEGE----- 346
QY 343 SREPSPIIANIEAVANSMGFKSHT 367
Db 347 ---EPSP-----EALVRYLSMRHT 363

```

RESULT 7

```

US-10-354-358-92
; Sequence 92, Application US/10354358
; Publication No. US20030157082A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc
; APPLICANT: Hunter, John Joseph
; APPLICANT: MacBeth, Kyle J.
; APPLICANT: Tsai, Fong-Ying
; APPLICANT: Lesoon, Andrea
; APPLICANT: Lightcap, Eric S.
; APPLICANT: Williamson, Mark
; APPLICANT: Rudolph-Owen, Laura A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING
; TITLE OF INVENTION: CANCER USING 140, 1470, 1586, 2089, 2427, 3702, 5891, 6428,
; TITLE OF INVENTION: 7181, 7660, 25641, 69583, 49863, 8897, 1682, 17667, 9235,
; TITLE OF INVENTION: 3703, 14171, 10359, 1680, 1450, 18894, 2088, 32427, 2160,
; TITLE OF INVENTION: 9252, 9389, 1642, 85269, 10297, 1584, 9525, 14124, 4469,
; TITLE OF INVENTION: 8990, 2100, 9288, 64698, 10480, 20893, 33230, 1586, 9943,
; TITLE OF INVENTION: 16334, 68862, 9011, 14031, 6178, 21225, 1420, 32236, 2099,
; FILE REFERENCE: MPI02-020PIRNONMIM
; CURRENT APPLICATION NUMBER: US/10/354,358
; PRIOR FILING DATE: 2003-01-30
; PRIOR APPLICATION NUMBER: US 60/353,600
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/364,517
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/371,075
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: US 60/371,507
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: US 60/372,984
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: US 60/374,194
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/382,995
; PRIOR FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/385,023
; PRIOR FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 60/388,853
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/389,395
; PRIOR FILING DATE: 2002-06-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: Fastseq for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 778
; TYPE: PRT
; ORGANISM: Homo sapiens

```

US-10-354-358-92

```

Query Match 26.3%; Score 602; DB 12; Length 778;
Best Local Similarity 42.8%; Pred. No. 2.9e-44;
Matches 110; Conservative 63; Mismatches 82; Indels 2; Gaps 2;
QY 8 VGKYEVRITIGEGTFAKVKFARNTDTGDNVAIKMAKSTILKNRMVDQIKREISIMKIVR 67
Db 31 VGPYLEKTLGKGQGLVKLVGHCHITGQKVAIKVNREK-LSESVMKVEREIALKLTIE 89
QY 68 HPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLSEESRKYFOQLVDAVAHCH 127
Db 90 HPNVLKLDHVDYENKKYLVLEHVSGGELFDYLVKKGRLTPKEARKFFRQIVSALDFCHS 149
QY 128 KGYYHRDLKPNLLDTNGNLKYSDFGLSALPOBQVHLLRTTCGTNYVAPEVLSGGYD 187
Db 150 YSICHRDLKPNLLDDEKNIRIADFGMASL-QVGSLLTSCGSPHYACPEVIKGEKYD 208
QY 188 GSAADIMSCGVILFVILAGVLPFSETDPLGLYRKINAABSCPPWFSAEVKFLIHRILD 247
Db 209 GRRADMWSCGVILFALLVGLALPFDNRLQLLEKVKRGVFMHFFIPPPDCQSLRGMI 268
QY 248 NPKTRIQIQIKKDPWF 264
Db 269 EPEKRLSLEQIQKHPWY 285

```

RESULT 8

```

US-10-116-326-2
; Sequence 2, Application US/10116326
; Publication No. US20030166889A1
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Brian
; APPLICANT: Priddle, Carl Johan
; TITLE OF INVENTION: No US20030166889A1el Human Kinases and Polynucleotides Encoding
; FILE REFERENCE: LEX-0332-USA
; CURRENT APPLICATION NUMBER: US/10/116,326
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/282,036
; PRIOR FILING DATE: 2001-04-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Fastseq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 778
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-116-326-2

```

```

Query Match 26.3%; Score 602; DB 12; Length 778;
Best Local Similarity 42.8%; Pred. No. 2.9e-44;
Matches 110; Conservative 63; Mismatches 82; Indels 2; Gaps 2;
QY 8 VGKYEVRITIGEGTFAKVKFARNTDTGDNVAIKMAKSTILKNRMVDQIKREISIMKIVR 67
Db 31 VGPYLEKTLGKGQGLVKLVGHCHITGQKVAIKVNREK-LSESVMKVEREIALKLTIE 89
QY 68 HPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLSEESRKYFOQLVDAVAHCH 127
Db 90 HPNVLKLDHVDYENKKYLVLEHVSGGELFDYLVKKGRLTPKEARKFFRQIVSALDFCHS 149
QY 128 KGYYHRDLKPNLLDTNGNLKYSDFGLSALPOBQVHLLRTTCGTNYVAPEVLSGGYD 187
Db 150 YSICHRDLKPNLLDDEKNIRIADFGMASL-QVGSLLTSCGSPHYACPEVIKGEKYD 208
QY 188 GSAADIMSCGVILFVILAGVLPFSETDPLGLYRKINAABSCPPWFSAEVKFLIHRILD 247
Db 209 GRRADMWSCGVILFALLVGLALPFDNRLQLLEKVKRGVFMHFFIPPPDCQSLRGMI 268
QY 248 NPKTRIQIQIKKDPWF 264
Db 269 EPEKRLSLEQIQKHPWY 285

```


US-10-354-358-8

Query Match 25.9%; Score 593.5; DB 12; Length 651;
Best Local Similarity 43.8%; Pred. No. 1.3e-43;
Matches 121; Conservative 58; Mismatches 92; Indels 5; Gaps 4;

QY 11 YEVGRITGEGTFAKVFARNTDGTGDNVAIKMASTILKKNRWVDQIKREISIMKIVRHPN 70
DB 11 YELHETIGTGFAKVKLACHILATGEMVAIKMDKTLGSD--LPRIKTEIETALKNLRHQH 68
QY 71 IVRLYEVLASPSKIYIIVLEFVTGGELFDRIVHKGRLEESRSKRYFOOLVDAVAHCHCKGV 130
DB 69 ICQLYHVLETKANKIFWLVLEPCGGELFDYIISQRLSEBETRVFRQIVSAVAVHSGQY 128
QY 131 YHRDLKPNLLDTNGNLKVSDFGLSALPOEGVEL-LRTTCGPNYVAPVLSGQGYDGS 189
DB 129 AHRDLKPNLLFDYHKLKIDFLGCAKPKGNKDYHLQTCGSLAYAAPELIQCKSVLGS 188
QY 190 AADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAREFSCPPWFSAEVKELIHRILDPNP 249
DB 189 EADVWSGILLYVLMCGFLPFDDNNVVALYKKIMRGKIDVFKWLSPPSILLLQMLQVDP 248
QY 250 KTRIQIGIKKDPWFRANY-VPIRAREEBE-VNLLD 283
DB 249 KKRISMKNLNHPWIMQDYNYPVWQSKNPFHLLD 284

RESULT 12

US-10-195-101-32
; Sequence 32, Application US/10195101
; Publication No. US20030087317A1
; GENERAL INFORMATION:
; APPLICANT: Bardman, Olga
; APPLICANT: Molteni, Angela
; APPLICANT: Magnaghi, Paola
; APPLICANT: Bosotti, Roberta
; APPLICANT: Scacheri, Emanuela
; APPLICANT: Isacchi, Antonella
; APPLICANT: Hodgson, David M.
; TITLE OF INVENTION: HUMAN NIM1 KINASE
; FILE REFERENCE: PC-0009-1 CIP
; CURRENT APPLICATION NUMBER: US/10/195,101
; PRIOR FILING DATE: 2002-07-11
; PRIOR FILING DATE: 2002-07-11
; PRIOR FILING DATE: 2002-07-11
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 793
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GenBank Accession No. US20030087317A1 g2052189
; DATABASE ENTRY DATE: 25 April 1997
US-10-195-101-32

Query Match 25.8%; Score 591; DB 15; Length 793;
Best Local Similarity 35.1%; Pred. No. 2.8e-43;
Matches 135; Conservative 72; Mismatches 134; Indels 44; Gaps 7;

QY 2 TKKRRVGVKVEGRITGEGTFAKVFARNTDGTGDNVAIKMASTILKKNRWVDQIKREIS 61
DB 51 TDSQPHIGNVRLQKTIGKNFAKVKARHVLTRGVAVKIIDK-TQINPTSQKLPRVR 109
QY 62 IMKIVRHPNIRVLYEVLASPSKIYIIVLEFVTGGELFDRIVHKGRLEESRSKRYFOOLVDA 121
DB 110 IMKILNHPNIVKLFEVETETKLYLVMYASGGEVFDYLVAGHMKKEARAKRQIVSA 169
QY 122 VAHCHCKGVYHRDLKPNLLDTNGNLKVSDFGLSALPOEGVELLRTTCGPNYVAPVLS 181
DB 170 VQYCHOKICVHRDLKPNLLDTNGNLKVSDFGLSALPOEGVELLRTTCGPNYVAPVLS 228

QY 182 SQQYDGSAAADIWSCGVILEVILAGYLPFSETDPLGLYRKINAAREFSCPPWFSAEVKFLI 241
DB 229 QGKKYDGPEDVWNSGLVILVTLVSGSLPFDQNLKELRERVLRGKYRVPYVMSTDCENLL 288
QY 242 HRILDPNPKTRIQIGIKKDPWFRANY-VPIRAREEBE-VNLLD 287
DB 289 KLLVLNPKTRIQIGIKKDPWFRANY-VPIRAREEBE-VNLLD 348
QY 288 ---FDGEGSYVA-----ENVNDEGFLMNAFEMITLSQNLNLSAL--- 327
DB 349 INDALYSQKVDYWAYVYILLGRKYPPEFEGGESLSG-----NLCQSRPSSDLNLSLQSP 404
QY 328 ---FDRQDFVKKRQTRFVSRREPS 348
DB 405 AHLKVQSRISANQKORRFSDHAGPS 429

RESULT 13

US-09-801-368-152
; Sequence 152, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR FILING DATE: 2001-03-07
; PRIOR FILING DATE: 2000-01-19
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 152
; LENGTH: 1518
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-152

Query Match 25.7%; Score 589.5; DB 10; Length 1518;
Best Local Similarity 30.7%; Pred. No. 8.9e-43;
Matches 138; Conservative 90; Mismatches 135; Indels 87; Gaps 10;

QY 2 TKKRR-----VGKVEVGRITGEGTFAKVFARNTDGTGDNVAIKMAKSTIL----- 48
DB 67 TKSKRKSRSDTVGPKWKLGTGKSSGRVRLAKRNETGQAAIKIVPKKAFVHCNNGT 126
QY 49 -----KNRMVDQ-----IKREISIMKIVRHPNIRVLYEVLASP 81
DB 127 VPSNYSSTSSNVTNVSPTSASREHNHNSQTPYGEREIVNKLSTNVMVAFVWENK 186
QY 82 SKIYTLVFTVGTGELFDRIVHKGRLEESRSKRYFOOLVDAVAHCHCKGVYHRDLKPNLL 141
DB 187 SELYLVLVVDGELFDYLVSKGLPEREAHYFQIVGVGSYCHSFNICHRLDLPENLL 246
QY 142 LD-TGNLKVSDPGLSALPOEGVELLRTTCGPNYVAPVLSGQGYDGSAAADIWSCGVIL 200
DB 247 LDKNRRIKIADFGMAALEPN-KLLKTCGSPHYASPEIVMGRPYHGGPSDWSGCVIL 305
QY 201 FVILAGYLPFSETDPLGLYRKINAAREFSCPPWFSAEVKELIHRILDPNPKTRIQIGIKK 260

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 16:50:02 ; Search time 20 Seconds
(without alignments)
943.532 Million cell updates/sec

Title: US-09-824-735-2
Perfect score: 2293
Sequence: 1 MTKKMRRVGKVEGRTIGEG.....IWRATEGIPKSEILRTITP 446

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/6C.COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	663	28.9	633	3	US-08-557-006C-43
2	662	28.9	552	3	US-08-557-006C-40
3	648	28.3	257	2	US-07-857-224B-25
4	643	28.0	345	3	US-09-101-146-1
5	636.5	27.8	257	3	US-09-101-146-6
6	591	25.8	793	4	US-08-523-849-32
7	584	25.5	729	2	US-08-677-298-2
8	584	25.5	729	4	US-09-523-849-33
9	579.5	25.3	779	4	US-08-817-832B-31
10	576.5	25.1	776	4	US-09-523-849-34
11	575	25.1	722	4	US-09-984-890-4
12	575	25.1	722	4	US-08-817-832B-32
13	575	25.1	745	4	US-09-523-849-36
14	574	25.0	668	4	US-09-930-181-2
15	574	25.0	724	4	US-09-984-890-2
16	561	24.5	604	4	US-09-523-849-35
17	543	23.7	339	3	US-08-688-988-33
18	523.5	22.8	351	3	US-08-688-988-28
19	511.5	22.3	631	4	US-09-579-664B-11
20	506	22.1	476	3	US-09-522-800-16
21	506	22.1	476	3	US-08-924-183-1
22	506	22.1	476	4	US-09-488-364-1
23	506	22.1	476	4	US-09-113-785-1
24	505.5	22.0	353	3	US-08-688-988-31
25	505.5	22.0	359	3	US-08-688-988-32
26	503.5	22.0	334	4	US-09-523-849-31
27	500.5	21.8	354	3	US-08-688-988-29

ALIGNMENTS

RESULT 1

US-08-557-006C-43
; Sequence 43, Application US/08557006C
; Patent No. 6258547
; GENERAL INFORMATION:

; APPLICANT: Beri, Rajindar K.
; APPLICANT: Carling, David
; APPLICANT: Forster, Robert A.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING AMP-ACTIVATED PROTEIN KINASE
; FILE REFERENCE: NGAP/PHM37588/UST
; CURRENT APPLICATION NUMBER: US/08/557,006C
; CURRENT FILING DATE: 1996-03-06
; PRIOR APPLICATION NUMBER: PCT/GB94/01093
; PRIOR FILING DATE: 1994-05-20
; PRIOR APPLICATION NUMBER: GB 9310489.1
; PRIOR FILING DATE: 1993-05-21
; PRIOR APPLICATION NUMBER: GB 9318010.7
; PRIOR FILING DATE: 1993-08-31
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 43
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Yeast
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(633)
; OTHER INFORMATION: Yeast SNF1 polypeptide
US-08-557-006C-43

Query Match 28.9%; Score 663; DB 3; Length 633;
Best Local Similarity 39.4%; Pred. No. 1.7e-58;
Matches 135; Conservative 80; Mismatches 98; Indels 30; Gaps 6;

QY	8	VGYEYVGRITGEGTFAKVFARNTDTCNDVAIKIMAKSTILKNRMVDQIKREISIMKIVR	67
DB	52	IGNYQIVKTLTGESFGKVLAYHTTTGQKVALINKVKLAKSDMQGREREISYLRLLR	111
QY	68	HPNIVLYEVLASPSKIYIVLEFVVTGGELFDRIHVHGRLEESERKYFQOLVDAVACHC	127
DB	112	HPHIIKLYDIKSDIEIMVIEY-AGNELFDYIVQDKMSEGEARFFQOIIISAVEYCHR	170
QY	128	KGVYHRDLKPENLLDNGNLKVSDFGLSALPQEGVELLRTTCGTENYVAPVLSQGVYD	187
DB	171	HKIVHRDLKPENLLDHLNKNVIAIDFGLSNIMTDG-NFLKTCGSNYAAPEVISGKLYA	229
QY	188	GSAADIWSCGVILFVILAGLPPFSDTLPLGLYRKINAAEFPFPWFAEVKFLIHILDP	247
DB	230	GPEVDVWSCGVILYVLMCRRLPFDDDSIFVLFPKINISNGVYTLFPKFLSPGAAGLIKRMILV	289

```
QY 248 NPKTRIQIGIKDPWFRIN---YV---PIRAREEE-----VNL D 282
Db 290 NPLNRSIHQMDQWFKVLPYVLLPDLKPHPEEENNDKSGSSPDNDIDNLY 349
QY 283 DIRAVFDIGSYVAENVRNDEGL---MKNAFEMITLSQGL 322
Db 350 NILSSTMGYKDEIYESLESSEDTAFNEIRDAYLWLIKENS L 392

RESULT 2
US-08-557-006C-40
; Sequence 40, Application US/08557006C
; Patent No. 6258547
; GENERAL INFORMATION:
; APPLICANT: Beri, Rajinder K.
; APPLICANT: Carling, David A.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING AMP-ACTIVATED PROTEIN KINASE
; FILE REFERENCE: NGAP/PM37588/UST
; CURRENT APPLICATION NUMBER: US/08/557,006C
; CURRENT FILING DATE: 1996-03-06
; PRIOR APPLICATION NUMBER: PCT/GB94/01093
; PRIOR FILING DATE: 1994-05-20
; PRIOR APPLICATION NUMBER: GB 9310489.1
; PRIOR FILING DATE: 1993-05-21
; PRIOR APPLICATION NUMBER: GB 9318010.7
; PRIOR FILING DATE: 1993-08-31
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 40
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Rat
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1)..(1747)
; OTHER INFORMATION: Full length cDNA sequence fragment of Human AMPK -
; OTHER INFORMATION: fragment begins at nucleotide 24 and ends with
; OTHER INFORMATION: nucleotide 1765
US-08-557-006C-40

Query Match 28.9%; Score 662; DB 3; Length 552;
Best Local Similarity 36.4%; Pred. No. 1.7e-58;
Matches 149; Conservative 79; Mismatches 147; Indels 34; Gaps 7;

QY 1 MTKKMR-----RVGKVEGTGTCGCTFAKVKFARNITDGTGNVAIKTMAKSTILKNRMVDQ 55
Db 1 MAEKQHDGRVKIGHYVLGDTLGVGTGKVGKIGEHQLTGHKVAVKILNRQKIRSLDWGK 60
QY 56 IKREISIMKIVRHPNIVRYEVLASPSKIYIVLEFVTGGELPDRIVHKGRLEBESGRKYF 115
Db 61 IKREIQNLKLFRRPHIILKIYQVISTDTPFMWEYVSGGELFDYICKHGRVEVEARRLF 120
QY 116 QQLVDAVAHCHGVVHRDLKPENLLDNGNLKVSDFGLSALPQBGVELLRTTCGTPNY 175
Db 121 QQILSADYDCHRMVVRDLKPENLLDQMNAKIADPGLSNMMSG-BFLRTSCGSPNY 179
QY 176 VAPEVLSGGYDGSAAIDWCSGVILFVILAGYLPFSETLPGLYRKINAAEFCPPWFSA 235
Db 180 AAPEVLSGLYAGPEVDIWSGVILYALLCGTLFPDDEHVPTLFKKIRGVFIPEYLNR 239
QY 236 EVKFLIHRILDNPNTKTRIQIGIKDPWFRINVPPIRAREEE--EVNLDIDRAVFDGIG- 292
Db 240 SIATLLMHLQVDPPLKRAIKOIREHENFKQDLPVLPFDEDSYDANVIDDEAVEKVECK 299
QY 293 -----GSYVAENVRNDEGLMNAFEMITLSQGLNLSALPDRQDPVKRQTRFVSRREPS 348
Db 300 FECTESEVNSLYSGDPQQLAVAYHLI-----IDNRIMNQASEFYFLASSPPT 348
QY 349 EIIANTEAVANGSGFKSHTRNF-----KTR--LEGSLSSIKAGQLAV 387
Db 349 GSFMDMAMHIPPGLKPHPERMPPIADSPKARCPLDALNTTKPKSLAV 397
```

RESULT 3

```
US-07-857-224B-25
; Sequence 25, Application US/07857224B
; Patent No. 5958784
; GENERAL INFORMATION:
; APPLICANT: Benner, Steven A.
; TITLE OF INVENTION: Predicting Folded Structures of Proteins
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven A. Benner
; STREET: Hadlaubstrasse 151
; CITY: Zurich
; STATE: none
; COUNTRY: Switzerland
; ZIP: (note: this is an international post code) CH-8092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/857,224B
; FILING DATE: 03/25/92
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: none
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (International) 41 1 632 2830
; TELEFAX: (International) 41 1 262 2437
; TELEX: none
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 257
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION:
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; FEATURE: Protein kinase; Table 8 Column 28
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Hanks, S. K.
; AUTHORS: Hunter, T.
; TITLE: The protein kinase family
; JOURNAL: Science
; VOLUME: 241
; PAGES: 42-52
; DATE: 1988
; US-07-857-224B-25

Query Match 28.3%; Score 648; DB 2; Length 257;
Best Local Similarity 46.3%; Pred. No. 1.4e-57;
Matches 120; Conservative 62; Mismatches 75; Indels 2; Gaps 2;

QY 9 GKYEVRIGETGCTFAKVKFARNITDGTGNVAIKMAKSTILKNRMVDQIKREISIMKIVRH 68
Db 1 GNYQIVKTLGEGSGFKVLAHYHTTGTQKVALKIINKVLAKSDMQGRIEREISYLRLLRH 60
QY 69 PNIVRYEVLASPSKIYIVLEFVTGGELPDRIVHKGRLEBESGRKYFQOLVDVAHCHCK 128
Db 61 PHIIKLYDIVIKSDIEIMVIEY-AGNELFDYIVQRDKMSEQEARRFQOIISAVEYCHRH 119
QY 129 GYVHRDLKPENLLDNGNLKVSDFGLSALPQBGVELLRTTCGTPNYVAPEVLSGGYD 188
Db 120 KIVHDLKPENLLDHLNVLKIADFGLSNIMTDG-NFLKTCGSPNYAAPEVISKULYAG 178
QY 189 SAADTWSGCVILFVILAGYLPFSETLPGLYRKINAAEFCPPWFSAFVKELIHRILDPN 248
Db 179 PEVDWMSGCVILYVLMCRLLPFDDDESIPVLFPKINSGVYVTLPKFLSPGAAGLIKEMLVN 238
QY 249 PKTRIQIGIKDPWFRIN 267
```

Db 239 PLNRISHEIMQDDWFKVD 257

RESULT 4

```

US-09-101-146-1
; Sequence 1, Application US/09101146
; Patent No. 6124125
; GENERAL INFORMATION:
; APPLICANT: Dartmouth College, St. Vincent's Institute of
; APPLICANT: Medical Research, Kemp et al.
; TITLE OF INVENTION: No. 6124125el AMP Activated Protein Kinase
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jane Massey Licata, Esq.
; STREET: 66 E. Main Street
; CITY: Marlton
; STATE: NJ
; COUNTRY: USA

```

COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PC
OPERATING SYSTEM: WINDOWS 95
SOFTWARE: WORDPERFECT 6.0 FOR WINDOWS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/101,146
FILING DATE: October 7, 1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PN7450
FILING DATE: 8 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,357
REFERENCE/DOCKET NUMBER: DC-0050
TELECOMMUNICATION INFORMATION:
TELEPHONE: (856) 810-1515
TELEFAX: (856) 810-1454
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 345
TYPE: Amino acid
TOPOLOGY: Linear
US-09-101-146-1

RESULT 5

US-09-101-146-6
 ? Sequence 6, Application US/09101146
 ? Patent No. 6124125
 ? GENERAL INFORMATION:
 ? APPLICANT: Dartmouth College, St. Vincents Institute of
 ? APPLICANT: Medical Research, Kemp et al.
 ? TITLE OF INVENTION: No. 6124125el AMP Activated Protein Kinase
 ? NUMBER OF SEQUENCES: 64
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Jane Massey Licata, Esq.
 ? STREET: 66 E. Main Street
 ? CITY: Marlton
 ? STATE: NJ
 ? COUNTRY: USA
 ? ZIP: 08053
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
 ? COMPUTER: IBM PC
 ? OPERATING SYSTEM: WINDOWS 95
 ? SOFTWARE: WORDPERFECT 6.0 FOR WINDOWS
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/09/101,146
 ? FILING DATE: October 7, 1998
 ? CLASSIFICATION: 435
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: PN7450
 ? FILING DATE: 8 JAN 1996
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Jane Massey Licata
 ? REGISTRATION NUMBER: 32,257
 ? REFERENCE/DOCKET NUMBER: DC-0050
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (856) 810-1515
 ? TELEFAX: (856) 810-1454
 ? INFORMATION FOR SEQ ID NO: 6:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 257
 ? TYPE: Amino acid
 ? TOPOLOGY: Linear
 ? US-09-101-146-6

Query Match	27.8%;	Score	636.5;	DB	3;	Length	257;
Best local similarity	47.5%;	Pred. No.	2e-56;				
Matches	122;	Conservative	54;	Mismatches	80;	Indels	1;
Gaps	1;						
Qy	9	GKYEVGSTIGEGTFAKVKFARNTDGTGDNVAIKIMAKSTILKNRMVDOIKREISIMKI	VRH	68			
Db	1	GHVILGDTLGVGTGFKVKVKGKHELTHGHKAVAKILNRQIKRSLDVDGKIRREIQNLK	LFRR	60			
Qy	69	PNIVRIYEVLASPSKIYIVLEFVTGGELPDRIYVHKGLBESESEKYPQOOLVDVAH	CHCK	128			
Db	61	PHIILKQYVISTPSDIPMWMEYVSGGELFDYICKNGELDEKESRLEFQILSGVDY	CHRH	120			
Qy	129	GVYHRDLKPNLLDITNGNLKVSDFGLSALPQEGVELLRITCGTPNYVAPESVLS	GQGYDG	188			
Db	121	MVYHRDLKPNVLLDAHNAKIADFLSNMMSDG-EFLRTSCGSPENYAAPEVISGR	LYAG	179			
Qy	189	SADIMSCGVILFVILAGYLPSESTDLPGLYLRKINAAEFSCPPWFSAEVKFLH	RIIDPN	248			
Db	180	PEVDIWSGGVILFALIGCTLPFFDDHVTFLFKICDGIFFTPQYLPNSVILSLKH	NLQVD	239			
Qy	249	PKTRIQIGIKKQDPWR	265				
Db	240	PMKRATIKDIREHEWFK	256				

RESULTS

US-09-523-849-32
; Sequence 32, Application US/09523849
; Patent No. 6458561

```
/
/ GENERAL INFORMATION:
/ APPLICANT: Bandman, Olga
/ APPLICANT: Molteni, Angela
/ APPLICANT: Magnaghi, Paola
/ APPLICANT: Bosotti, Roberta
/ APPLICANT: Scacheri, Emanuela
/ APPLICANT: Isacchi, Antonella
/ APPLICANT: Hodgson, Dave
/ TITLE OF INVENTION: HUMAN NIM1 KINASE
/ FILE REFERENCE: PC-0009 US
/ CURRENT APPLICATION NUMBER: US/09/523,849
/ CURRENT FILING DATE: 2000-03-13
/ NUMBER OF SEQ ID NOS: 39
/ SOFTWARE: PERL Program
/ SEQ ID NO 32
/ LENGTH: 793
/ TYPE: PRT
/ ORGANISM: Rattus norvegicus
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: GenBank Accession No. 6458561 g2052189
US-09-523-849-32

Query Match      25.8%; Score 591; DB 4; Length 793;
Best Local Similarity 35.1%; Pred. No. 4.8e-51;
Matches 135; Conservative 72; Mismatches 134; Indels 44; Gaps 7;

QY 2 TKMRRVGVKEVGTGEGTFAKVKFARNVTDGDNVAIKIMAKSTILKNRMVDPQIKREIS 61
Db 51 TDEQPHIGNYRLQKTIGKGNFAKVKLARHVLTGREVAVKTIIDK-TQLNPTSLQKLFREVR 109

QY 62 IMKIVRPNVRLVYELVSLSPKIVYLVLEFVTGGELEPDRIVHKGLESESRKYFOQLVDA 121
Db 110 IMKILNPNVYKLVFEVETETKTYLWVEYASGGEVFDYLVAGHGMKEKEARAKPRQIVSA 169

QY 122 VACHCKGVVHRDLKPNLLDNGNLKVSDFGLSALPQGVGELLRTTCCTPNVVAPEVL 181
Db 170 VQYCHQKCIVHRDLKAENLLDADNMIKIADFGSNEFTVG-NKLDTFCSPPYAPAEFL 228

QY 182 SGQGYDGAADINSCGVILFVILAGVLPFSETDLPGLYRKINAAEFCPPWFSAEVLKFLI 241
Db 229 QGKKYDGPVDVWSLGVILYTLVSGSLPFDGQNLKELRERVLRGKYRVPVYMTDCENLL 288

QY 242 HRILDPNPKTRIQIGIKKDPWFRNLV--VPIRAREEEVNLDDIRAV----- 287
Db 289 KLLVLPNPKGSLGQIMKDRWNVGHEBELKPYSEPELDLDAKRIIDLVMTWGFARDE 348

QY 288 -----FDGIEGSYVA-----ENVERNDEGLPMNAFEMITLSQGLNLSAL--- 327
Db 349 INDALVSQKYDEVWATVILLGRKPPPEFGESLSG---NLCQRSEPSDLNNSLTQSP 404

QY 328 ----FRRQDFVKEOTRFFVSRREPS 348
Db 405 AHLKVORSISANOKORRFDHAGPS 429

RESULT 7
US-08-677-298-2
/ Sequence 2, Application US/08677298
/ Patent No. 5863729
/ GENERAL INFORMATION:
/ APPLICANT: Pivnicka-Worms, Helen
/ TITLE OF INVENTION: DNA SEQUENCES ENCODING HUMAN TCAR-1
/ TITLE OF INVENTION: KINASE
/ NUMBER OF SEQUENCES: 17
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
/ STREET: 5370 Manhattan Circle, Suite 201
/ CITY: Boulder
/ STATE: CO
/ COUNTRY: USA
/ ZIP: 80303
/ COMPUTER READABLE FORM:
```

```
/
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA: US/08/677,298
/ APPLICANT: Scacheri, Emanuela
/ FILING DATE: 09-JUL-1996
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Caruthers, Jennie M.
/ REGISTRATION NUMBER: 34,464
/ REFERENCE/DOCKET NUMBER: 9-96
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (303) 499-8080
/ TELEFAX: (303) 499-8089
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 729 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-677-298-2

Query Match      25.5%; Score 584; DB 2; Length 729;
Best Local Similarity 33.8%; Pred. No. 2.2e-50;
Matches 139; Conservative 81; Mismatches 154; Indels 40; Gaps 7;

QY 8 VGYKEVGRITGEGTFAKVKFARNVTDGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVR 67
Db 53 IGYRLKTKTIGKGNFAKVKLARHILTGREVAIKIIDK-TQLNPTSLQKLFREVRIMKILN 111

QY 68 HPMIVRLVYELVSLSPKIVYLVLEFVTGGELEPDRIVHKGLESESRKYFOQLVDAVHCHC 127
Db 112 HPMIVKLVFEVETETKTYLWVEYASGGEVFDYLVAGHGMKEKEARSKFRQIVSAVQYCHQ 171

QY 128 KGVYHRLDLPENLLDNGNLKVSDFGLSALPQGVGELLRTTCCTPNVVAPEVLSCGYD 187
Db 172 KRIVHRLDLPENLLDADNMIKIADFGSNEFTVGK-LDTFCGSPYAPAEFLQGGKYD 230

QY 188 GSAADINSCGVILFVILAGVLPFSETDLPGLYRKINAAEFCPPWFSAEVLKFLIHLRILDP 247
Db 231 GPEVDVWSLGVILYTLVSGSLPFDGQNLKELRERVLRGKYRIPFYMTDCENLLKRLVL 290

QY 248 NPKTRIQIGIKKDPWFRNLV--VPIRAREEEVNLDDIRAV----- 287
Db 291 NPKTRIQIGIKKDPWFRNLV--VPIRAREEEVNLDDIRAV----- 287

QY 288 ----FDGIEGSYV-----AENVERNDEGLPMNAFEMITLSQGLNLSA-----LFRROD 333
Db 351 KMKYDEITATYLLGRKSELSDSSSSNLSLAKVRPSSDLNNSLTQSPHKKVQSVS 410

QY 334 FVKRQTRFVSRREPSIIANIEAVANSMGFKSHTRNFKTRLEGLSSIKAGQLAV 387
Db 411 SSQKQRRYSDHAGP---AIPSVVAYPKESQISTADGLKEDGIGSRKSGSAV 460
```

```
RESULT 8
US-09-523-849-33
/ Sequence 33, Application US/09523849
/ Patent No. 6458561
/ GENERAL INFORMATION:
/ APPLICANT: Bandman, Olga
/ APPLICANT: Molteni, Angela
/ APPLICANT: Magnaghi, Paola
/ APPLICANT: Bosotti, Roberta
/ APPLICANT: Scacheri, Emanuela
/ APPLICANT: Isacchi, Antonella
/ APPLICANT: Hodgson, Dave
/ TITLE OF INVENTION: HUMAN NIM1 KINASE
/ FILE REFERENCE: PC-0009 US
/ CURRENT APPLICATION NUMBER: US/09/523,849
/ CURRENT FILING DATE: 2000-03-13
/ NUMBER OF SEQ ID NOS: 39
```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 779 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-817-832B-31
Query Match      25.3%; Score 579.5; DB 4; Length 779;
Best Local Similarity 35.2%; Pred No. 6.9e-50;
Matches 136; Conservative 70; Mismatches 135; Indels 45; Gaps 8
QY   2 TKORRVGYEVRTIGEGTFAKVFARNTDGDVNAIKMAKSTILKNRMVDQIKREIS 61
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   62 IMKIIVHPIVLYELASPSKIYIIVLEFVTGGELFDRIVHKGRLEESSESKYFOOLVDA 121
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   96 IMKILNHPINVLFEVETEKTLYLMEYASGEVDFYLVAHGRMKEAKRQIVSA 155
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   122 VAHCCKGVYHRDLKPENLLDTNGNLKVSDFGLSALPQEGVELLRTTCGTPNVVAPEVL 181
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   156 VOYCHOKCIHVHRLDKAENLLLDADWNKIADFGFSNETVG-NKLDTFCGSPFYAAPELF 214
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   182 SQOYGDSAADISWCGVILFVILAGYLPFSETDPLGL-YRKINAAEFCPPWFSAEVKFL 240
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   215 QGKKYDGPEDVWSLGVIYLTVSGSLFPDGQNLKELRERSCLRGKYRVPFYMSTDCE 274
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   241 IIRILLDPNEKTIOIQGIKKDPWFLNY--VIPAREEEVNLDIDRAV----- 287
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   275 LKKLLVLNPIKGSLEQIMKORMNVGHHEBELKPYSEPELDNDAKEIDIMVTMGFARD 334
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   288 -----PDGIBGSYVA-----ENVRNDEGPLMNAPEMITLSQGLNISAL-- 327
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   335 EINDALVSKQYDEVMTATVILLGRKPEPFEGGESISSG----NLCORSPPSSDLANNSTLQS 390
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   328 -----FDRQRDFVKRQTRFVSRREPS 348
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   391 PAHLKVORTISANQKRFFSDHAGPS 416
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 10
US-09-523-849-34
; Sequence 34, Application US/09523849
; Patent No. 6458561
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Molteni, Angela
; APPLICANT: Magnaghi, Paola
; APPLICANT: Bosotti, Roberta
; APPLICANT: Scacheri, Emanuela
; APPLICANT: Isacchi, Antonella
; APPLICANT: Hodgson, Dave
; TITLE OF INVENTION: HUMAN NIMI KINASE
; FILE REFERENCE: PC-0009 US
; CURRENT APPLICATION NUMBER: US/09/523/849
; CURRENT FILING DATE: 2000-03-13
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PERL Program
; SEQ ID NO 34
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GenBank Accession No. 6458561 95672676
; US-09-523-849-34
Query Match      25.1%; Score 576.5; DB 4; Length 776;
Best Local Similarity 34.3%; Pred. No. 1.4e-49;
Matches 146; Conservative 78; Mismatches 147; Indels 55; Gaps 11;
QY   3 KMORVGYEVRTIGEGTFAKVFARNTDGDVNAIKMAKSTILKNRMVDQIKREISI 62
```

```

Db      19  QKPLRVGYDVERILGKGFVAVKLARHRTKTQVAIKLIDK-TLDSNLEKIYREVQL 77
QY      63  MKIVRHPNIVLYEVLASPSKIYIVLEFVTGGLFDRIVHKGRLSEESRKRYFOQLVDVAV 122
Db      78  MKLLNHPNIIKLYQVMEKDMLYIVTEPAKNGEMFDYLTSGHLSENEARKKFWQLSAV 137
QY      123  AHCHCKGVYHRDLKPNLLDNGNLKYSDFGLSALPOEGVELLRTTCGTNYVAPEVLS 182
Db      138  EYCHNHIVHRDLKTNLGLDNGMDIKLADFGNGFYKPG-EPLSTWCGSPPYAAPEVFE 196
QY      183  GQGYDGAADWSCGVILFVILAGYLPSETDLGLYRKINAAEFSCPPWPSAEVKFLIH 242
Db      197  GKEYEGQLDLSGLVLYLVCSGLPDGPNLPTLQORVLEGRFPFPFMSQCEILIR 256
QY      243  RILDNPKTRIQIOGIKKDPWFRNLNYPVIRAREBEV-----NLDD-----IRA 286
Db      257  RMLVVDPAKRITIAQIRQHRWQAD--PTLLQODDPAFSGMQYTSNLGDYNEQVGLIMQA 314
QY      287  VFDCIGSYVAENVERNDEGLMNAPEMITLSQGLNLSALFDRRQDFVKQTRFVSRR 346
Db      315  L--GIDQRTVESQNS-----YNHFAAI-----YLLERLRE--HRSTOPSSRAT 358
QY      347  PSEIANIEAVANSMGFKSHTRNFKTRLEGSLSSIKAGQLAVIYIYEVAPSLFVMDVRKA 406
Db      359  PAP-----ARQPLANSLSLEVPQELPCDPPR--PSLLCPQPAL 399
QY      407  AGETLE 412
Db      400  AQSVLQ 405

```

```

RESULT 11
US-09-884-890-4
; Sequence 4, Application US/09984890
; Patent No. 6492156
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001306
; CURRENT APPLICATION NUMBER: US/09/984.890
; CURRENT FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 722
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-884-890-4

Query Match      25.1%; Score 575; DB 4; Length 722;
Best Local Similarity 41.9%; Pred. No. 1.8e-49;
Matches 114; Conservative 61; Mismatches 89; Indels 8; Gaps 3;

QY      8  VGKYEVRTIGEGTFAKVKFARNITDGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVR 67
Db      50  IGYRLTKTGKGNFAKVKLARHLLTGKEVAVKIIDK-TQLNSSSLQKLFREVRIMKVLN 108
QY      68  HPNIVLYEVLASPSKIYIVLEFVTGGLFDRIVHKGRLSEESRKRYFOQLVDVAHCHC 127
Db      109  HPNIVKLFEVIEETKTLVMEYASGGEVDFLVAGHMKKEARAKFRQIVSAVQYCHQ 168
QY      128  KGIVHRDLKPNLLDNGNLKYSDFGLSALPOEGVELLRTTCGTNYVAPEVLSGQGYD 187
Db      169  KFIVHRDLKAENLLDADNMNIKIADFGSNEFTFG-NKLDTCGSPPYAAPELFOGKKYD 227
QY      188  GSAADIWSCGVILFVILAGYLPSETDLGLYRKINAAEFSCPPWPSAEVKFLIHRILD 247
Db      228  GPEVDVWSLGVILYTLVSGSLPFDGQNLKELREVRILGKYRIPFYMSTDCENLLKKFLIL 287
QY      248  NPKTRIQIOGIKKDPWFRNLNYPVIRAREBEV 279
Db      288  NPSKRGTLQIMKDRWMNVGH-----EDEL 313

RESULT 13
US-09-523-849-36
; Sequence 36, Application US/09523849
; Patent No. 6458561
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga

```

```

Db      288  NPSKRGTLQIMKDRWMNVGH-----EDEL 313

RESULT 12
US-08-817-832B-32
; Sequence 32, Application US/08817832B
; Patent No. 6579691
; GENERAL INFORMATION:
; APPLICANT: MANDELKOW, Eckhard, et al.
; TITLE OF INVENTION: No. 6579691el Protein Kinase (NPK-110)
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 S. Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: US
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA: US/08/817,832B
; APPLICATION NUMBER: WO PCT/EP95/04258
; FILING DATE: 28-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 30-OCT-1995
; PRIOR APPLICATION DATA: EP 94 11 7122.5
; FILING DATE: 28-OCT-1994
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 722 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-817-832B-32

```

```

Query Match      25.1%; Score 575; DB 4; Length 722;
Best Local Similarity 41.9%; Pred. No. 1.8e-49;
Matches 114; Conservative 61; Mismatches 89; Indels 8; Gaps 3;

QY      8  VGKYEVRTIGEGTFAKVKFARNITDGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVR 67
Db      50  IGYRLTKTGKGNFAKVKLARHLLTGKEVAVKIIDK-TQLNSSSLQKLFREVRIMKVLN 108
QY      68  HPNIVLYEVLASPSKIYIVLEFVTGGLFDRIVHKGRLSEESRKRYFOQLVDVAHCHC 127
Db      109  HPNIVKLFEVIEETKTLVMEYASGGEVDFLVAGHMKKEARAKFRQIVSAVQYCHH 168
QY      128  KGIVHRDLKPNLLDNGNLKYSDFGLSALPOEGVELLRTTCGTNYVAPEVLSGQGYD 187
Db      169  KFIVHRDLKAENLLDADNMNIKIADFGSNEFTFG-NKLDTCGSPPYAAPELFOGKKYD 227
QY      188  GSAADIWSCGVILFVILAGYLPSETDLGLYRKINAAEFSCPPWPSAEVKFLIHRILD 247
Db      228  GPEVDVWSLGVILYTLVSGSLPFDGQNLKELREVRILGKYRIPFYMSTDCENLLKKFLIL 287
QY      248  NPKTRIQIOGIKKDPWFRNLNYPVIRAREBEV 279
Db      288  NPSKRGTLQIMKDRWMNVGH-----EDEL 313

```

```

RESULT 13
US-09-523-849-36
; Sequence 36, Application US/09523849
; Patent No. 6458561
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga

```

```

; APPLICANT: Molteni, Angela
; APPLICANT: Magnaghi, Paola
; APPLICANT: Bosotti, Roberta
; APPLICANT: Scaccheri, Emanuela
; APPLICANT: Isacchi, Antonella
; APPLICANT: Hodgson, Dave
; TITLE OF INVENTION: HUMAN NIM1 KINASE
; FILE REFERENCE: PC-0009 US
; CURRENT APPLICATION NUMBER: US/09/523,849
; CURRENT FILING DATE: 2000-03-13
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PERL Program
; SEQ ID NO 36
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: GenBank Accession No. G1749794
US-09-523-849-36

Query Match      25.1%; Score 575; DB 4; Length 745;
Best Local Similarity 41.9%; Pred. No. 1.e-49;
Matches 114; Conservative 61; Mismatches 89; Indels 8; Gaps 3;

QY   8 VGKYEVRTIGEGTFAKVAFARNTDTGDNVAIKMAKSTILKNRMVDOIKREISIMKIIVR 67
DB   17 IGNRYLLKTIGKNGFAVKVLARHILTCKEVAVKIIDK-TQLNSSSLQKLFRVIRMKVLN 75
QY   68 HPNIIVLYEVLASPSKIYIVLEFVTGGELFDRIIVKHGRLESESRKYFOQLVDVAHCHC 127
DB   76 HPNIIVKLFVEIETETKTYLVMEYASGGGEVDYLVAHGMRKEKEAKRFQIVSAVQYCHQ 135
QY   128 KGVIHRDLKPENILLDTNGNLKVSDFGLSALPQEGVELLRITTCGTNNYVAPEVLSQGQYD 187
DB   136 KFIVHRDLKAENLLLDADNMNKIAIDFGFSNEFTTG-NKLDTFCGSPPYAAPELFGKKYD 194
QY   188 GSAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFCPPWFSAEVKFLIHRILD 247
DB   195 GPEDVMSLGVLITLVSGSLFFDGQNKLKERVLRGKRIPIPFYMSTDCEMLKKFLIL 254
QY   248 NPKTRIQIGIKDPWFRNLNYVIPRARBEVV 279
DB   255 NPSKRGTLEQIMKDRMNVGH-----EDEDEL 280

RESULT 14
US-09-330-181-2
; Sequence 2, Application US/09930181
; Patent No. 6455292
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Full-length Serine Protein Kinase in Brain and Pancreas
; FILE REFERENCE: 16U 101 V1
; CURRENT APPLICATION NUMBER: US/09/930,181
; CURRENT FILING DATE: 2001-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-930-181-2

Query Match      25.0%; Score 574; DB 4; Length 668;
Best Local Similarity 41.2%; Pred. No. 2e-49;
Matches 106; Conservative 64; Mismatches 85; Indels 2; Gaps 2;

QY   8 VGKYEVRTIGEGTFAKVAFARNTDTGDNVAIKMAKSTILKNRMVDOIKREISIMKIIVR 67
DB   16 VGPRLKTLKGQTGLVKLGVCHVCCTCKVAKIWNREK-LSESVMKVRERIALKLIIE 74
QY   68 HPNIIVLYEVLASPSKIYIVLEFVTGGELFDRIIVKHGRLESESRKYFOQLVDVAHCHC 127

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 16:45:11 ; Search time 44 Seconds
(without alignments)
1608.910 Million cell updates/sec

Title: US-09-824-735-2

Perfect score: 2293

Sequence: 1 MTKKRRVGKYEVRTIGEG.....IINPATEGIPKSEILRTITF 446

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq 19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2293	100.0	446	21	AA050962
2	2293	100.0	446	23	ABG31356
3	2288	99.8	446	23	ABG31357
4	2285	99.7	446	23	ABG31358
5	2273	99.1	442	21	AA050963
6	2085	90.9	405	21	AA050964
7	1484.5	64.7	445	21	AA051348
8	1484.5	64.7	445	24	AB040808
9	1469	64.1	481	21	AA034115

10	1469	64.1	481	24	ABR40707	Zea mays oil trait
11	1367	59.6	441	21	AA03423	Soybean putative c
12	1367	59.6	441	24	ABR40717	Glycine max oil tr
13	1344.5	58.6	416	21	AA014342	Arabidopsis thalia
14	1336.5	58.3	409	21	ABG31349	Arabidopsis thalia
15	1314	57.3	441	24	ABR40817	Arabidopsis thalia
16	1303.5	56.8	396	21	AA031350	P. patens protein
17	1288.5	56.2	441	22	AA014343	Arabidopsis thalia
18	1136.5	52.2	380	21	AA014343	Arabidopsis thalia
19	1163.5	50.7	367	21	AA014344	Arabidopsis thalia
20	1063	46.4	422	21	AA034119	Soybean putative c
21	1063	46.4	422	24	ABR40713	Glycine max oil tr
22	1043	45.5	438	21	AA03421	Soybean putative c
23	1043	45.5	438	24	ABR40715	Glycine max oil tr
24	1040	45.4	441	24	ABR40814	Arabidopsis thalia
25	1039.5	45.3	431	21	AA039524	Arabidopsis thalia
26	1039.5	45.3	439	21	AA039523	Arabidopsis thalia
27	997.5	43.5	332	21	AA052661	Arabidopsis thalia
28	970.5	42.3	332	21	AA052662	Arabidopsis thalia
29	970.5	42.3	374	21	AA034603	Arabidopsis thalia
30	932	40.6	380	21	AA026367	Arabidopsis thalia
31	930	40.6	366	21	AA026368	Arabidopsis thalia
32	911	39.7	416	21	AA066639	Arabidopsis thalia
33	911	39.7	433	21	AA066638	Arabidopsis thalia
34	876.5	38.2	376	21	AA039525	Arabidopsis thalia
35	849.5	37.0	273	21	AA052663	Arabidopsis thalia
36	849.5	37.0	335	21	AA034604	Arabidopsis thalia
37	835	36.4	351	21	AA049615	Arabidopsis thalia
38	833	36.3	337	21	AA049616	Arabidopsis thalia
39	809.5	35.3	324	21	AA034605	Arabidopsis thalia
40	808.5	35.3	371	21	AA054418	Zea mays protein f
41	808.5	35.3	431	21	AA048083	Arabidopsis thalia
42	808.5	35.3	455	21	AA048082	Arabidopsis thalia
43	775	33.8	307	21	AA026369	Arabidopsis thalia
44	757	33.0	353	21	AA067000	Arabidopsis thalia
45	719	31.4	523	21	AA03425	Wheat putative car

ALIGNMENTS

RESULT 1
AAG50962
ID AAG50962 standard; Protein; 446 AA.
XX AAG50962;
XX AAG50962;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 64634.
XX Arabidopsis thaliana protein fragment SEQ ID NO: 64634.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX Arabidopsis thaliana.
XX EF1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.


```

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 2293; DB 21; Length 446;
Best Local Similarity 100.0%; Pred. No. 5e-237;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKKRRVGKYEVRTIGEGTFAKVFARNNTDGDVAIKIMAKSTILKNRWVDQIKREI 60
DB 1 MTKKRRVGKYEVRTIGEGTFAKVFARNNTDGDVAIKIMAKSTILKNRWVDQIKREI 60

QY 61 SIMKIVRHNPVRLVLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLESESRKYFQQLVD 120
DB 61 SIMKIVRHNPVRLVLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLESESRKYFQQLVD 120

QY 121 AVAHCCKGVYHRDLKPENLDDTGNLKVSDFGLSALPQEGVELLRTTCGTNYVAPEV 180
DB 121 AVAHCCKGVYHRDLKPENLDDTGNLKVSDFGLSALPQEGVELLRTTCGTNYVAPEV 180

QY 181 LSGQYDGSAAADWSGVLFFVILAGYLPFSETDPLGLYRKINAEPSCPPWFSAEVKFL 240
DB 181 LSGQYDGSAAADWSGVLFFVILAGYLPFSETDPLGLYRKINAEPSCPPWFSAEVKFL 240

QY 241 IHRILDPNPKTRIQIGIKKDPWFLNYPVIRAREEEVNLDIDRAVDFGIEGSAVAENV 300
DB 241 IHRILDPNPKTRIQIGIKKDPWFLNYPVIRAREEEVNLDIDRAVDFGIEGSAVAENV 300

QY 301 ERNDEGLMNAFEMITLSQGLNLSALFDRRQDFVKQTRFVSRREPSIIANIEAVNS 360
DB 301 ERNDEGLMNAFEMITLSQGLNLSALFDRRQDFVKQTRFVSRREPSIIANIEAVNS 360

QY 361 MGFKSHTRNFKTRLEGLSSIKAGQAVVIEIYEVAPSLFMDVVRKAAGETLVEYHKFYKKL 420
DB 361 MGFKSHTRNFKTRLEGLSSIKAGQAVVIEIYEVAPSLFMDVVRKAAGETLVEYHKFYKKL 420

QY 421 CSKLENIWEATGIPKSEILRTITF 446
DB 421 CSKLENIWEATGIPKSEILRTITF 446

RESULT 2
ID ABG31356
XX ABG31356 standard; Protein; 446 AA.
AC ABG31356;
XX
XX
DT 15-NOV-2002 (first entry)
XX
DE Arabidopsis thaliana SOS2 serine/threonine protein kinase.
XX
XX Salt overly sensitive 7; SOS2; serine/threonine protein kinase;

```

```

KW salt tolerance; agricultural crop; rice; corn; wheat; cotton;
KW peanut; soybean; plant protectant; plant; transgenic.
OS Arabidopsis thaliana.
XX US2002095032-A1.
XX 18-JUL-2002.
PD
XX
XX 04-APR-2001; 2001US-0824735.
PF
XX
XX 04-APR-2000; 2000US-194649P.
PR
XX
XX (UYAR-) UNIV ARIZONA.
PA
XX
XX Zhu J, Liu J, Ishitani M, Kim C, Halfter U;
PI
XX
XX WPI; 2002-665801/71.
XX N-PSDB; ABK91073, ABK91084.
DR
XX
XX New SOS2 protein kinase polypeptide and nucleic acids encoding the
PT polypeptide, useful for increasing salt tolerance to plants, especially
PT to agricultural crops -
XX
XX
XX Claim 37; Fig 2; 28pp; English.
XX
XX The present invention relates to the isolation of an Arabidopsis
CC thaliana gene encoding salt overly sensitive 7 (SOS2) serine/threonine
CC protein kinase. The polynucleotide sequence encoding SOS2 can be used
CC to increase salt tolerance in plants, particularly agricultural crops
CC such as rice, corn, wheat, cotton, peanut and soybean. The
CC polynucleotide sequence encoding SOS2 can also be used to generate
CC transgenic plants, as hybridization probes for RNA, cDNA and DNA to
CC isolate those cDNAs or genes which exhibit a high degree of similarity
CC to the sequence of the SOS2 gene, and as PCR primers for the production
CC of DNA which encodes an enzyme having serine/threonine protein kinase
CC activity. The present sequence represents Arabidopsis thaliana SOS2
CC serine/threonine protein kinase.
XX
XX Sequence 446 AA;

```

```

Query Match 100.0%; Score 2293; DB 23; Length 446;
Best Local Similarity 100.0%; Pred. No. 5e-237;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKKRRVGKYEVRTIGEGTFAKVFARNNTDGDVAIKIMAKSTILKNRWVDQIKREI 60
DB 1 MTKKRRVGKYEVRTIGEGTFAKVFARNNTDGDVAIKIMAKSTILKNRWVDQIKREI 60

QY 61 SIMKIVRHNPVRLVLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLESESRKYFQQLVD 120
DB 61 SIMKIVRHNPVRLVLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLESESRKYFQQLVD 120

QY 121 AVAHCCKGVYHRDLKPENLDDTGNLKVSDFGLSALPQEGVELLRTTCGTNYVAPEV 180
DB 121 AVAHCCKGVYHRDLKPENLDDTGNLKVSDFGLSALPQEGVELLRTTCGTNYVAPEV 180

QY 181 LSGQYDGSAAADWSGVLFFVILAGYLPFSETDPLGLYRKINAEPSCPPWFSAEVKFL 240
DB 181 LSGQYDGSAAADWSGVLFFVILAGYLPFSETDPLGLYRKINAEPSCPPWFSAEVKFL 240

QY 241 IHRILDPNPKTRIQIGIKKDPWFLNYPVIRAREEEVNLDIDRAVDFGIEGSAVAENV 300
DB 241 IHRILDPNPKTRIQIGIKKDPWFLNYPVIRAREEEVNLDIDRAVDFGIEGSAVAENV 300

QY 301 ERNDEGLMNAFEMITLSQGLNLSALFDRRQDFVKQTRFVSRREPSIIANIEAVNS 360
DB 301 ERNDEGLMNAFEMITLSQGLNLSALFDRRQDFVKQTRFVSRREPSIIANIEAVNS 360

QY 361 MGFKSHTRNFKTRLEGLSSIKAGQAVVIEIYEVAPSLFMDVVRKAAGETLVEYHKFYKKL 420
DB 361 MGFKSHTRNFKTRLEGLSSIKAGQAVVIEIYEVAPSLFMDVVRKAAGETLVEYHKFYKKL 420

```

```

QY 421 CSKLENIWRATEGIPKSEILRTITF 446
DB 421 CSKLENIWRATEGIPKSEILRTITF 446

RESULT 3
ID ABG31357 standard; Protein; 446 AA.
AC ABG31357;
XX
XX 15-NOV-2002 (first entry)
DT
XX Arabidopsis thaliana SOS2 mutant (K40N) serine/threonine protein kinase.
DE
XX Salt overly sensitive 7; SOS2; serine/threonine protein kinase;
KW salt tolerance; agricultural crop; rice; corn; wheat; cotton;
KW peanut; soybean; plant protectant; plant; transgenic; mutant; mutein.
XX
XX Arabidopsis thaliana.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Misc-difference 40
FT /note= "Substitution of wild type Lys to Asn"
FT
XX
XX US2002095032-A1.
XX
XX 18-JUL-2002.
XX
XX 04-APR-2001; 2001US-0824735.
XX
XX 04-APR-2000; 2000US-194649P.
XX
XX (UYAR-) UNIV ARIZONA.
XX
XX Zhu J, Liu J, Ishitani M, Kim C, Halfter U;
XX WPI; 2002-665801/71.
XX
XX New SOS2 protein kinase polypeptide and nucleic acids encoding the
PT polypeptide, useful for increasing salt tolerance to plants, especially
PT to agricultural crops -
XX
XX Disclosure; Page -; 28pp; English.
XX
XX The present invention relates to the isolation of an Arabidopsis
CC thaliana gene encoding salt overly sensitive 7 (SOS2) serine/threonine
CC protein kinase. The polynucleotide sequence encoding SOS2 can be used
CC to increase salt tolerance in plants, particularly agricultural crops
CC such as rice, corn, wheat, cotton, peanut and soybean. The
CC polynucleotide sequence encoding SOS2 can also be used to generate
CC transgenic plants, as hybridization probes for RNA, cDNA and DNA to
CC isolate those cDNAs or genes which exhibit a high degree of similarity
CC to the sequence of the SOS2 gene, and as PCR primers for the production
CC of DNA which encodes an enzyme having serine/threonine protein kinase
CC activity. The present sequence represents a mutant of Arabidopsis
CC thaliana SOS2 serine/threonine protein kinase.
CC Note: The present sequence is not given in the specification but
CC is created by the indexer from the information given on page 5.
XX
XX Sequence 446 AA;

Query Match 99.8%; Score 2288; DB 23; Length 446;
Best Local Similarity 99.8%; Pred. No. 1.7e-236;
Matches 445; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKKMRVGKYEYGRITIGETFAKVFARNITDGDVNAIKMAKSTILKNRWVDQIKREI 60
DB 1 MTKKMRVGKYEYGRITIGETFAKVFARNITDGDVNAIKMAKSTILKNRWVDQIKREI 60
QY 61 SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLSESRKYFQOLVD 120

```

CC polynucleotide sequence encoding SOS2 can also be used to generate
 CC transgenic plants, as hybridization probes for RNA, cDNA and DNA to
 CC isolate those cDNAs or genes which exhibit a high degree of similarity
 CC to the sequence of the SOS2 gene, and as PCR primers for the production
 CC of DNA which encodes an enzyme having serine/threonine protein kinase
 CC activity. The present sequence represents a mutant of Arabidopsis
 CC thaliana SOS2 serine/threonine protein kinase.
 CC Note: The present sequence is not given in the specification but
 CC is created by the indexer from the information given on page 5.
 XX

SQ Sequence 446 AA;

Query Match 99.7%; Score 2285; DB 23; Length 446;
 Best Local Similarity 99.8%; Pred. No. 3.6e-236;
 Matches 445; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKMRRVGKYEVRTIGEGTFAKVFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREI 60
 DB 1 MTKMRRVGKYEVRTIGEGTFAKVFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREI 60

QY 61 SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKVFQQLVD 120
 DB 61 SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKVFQQLVD 120

QY 121 AVAHCCKGVYHRDLKPENLILLDTGNLKVSDPGLSALPOBQVLLRTTCGTPNYVAPEV 180
 DB 121 AVAHCCKGVYHRDLKPENLILLDTGNLKVSDPGLSALPOBQVLLRTTCGTPNYVAPEV 180

QY 181 LSGQGYDGSAAIDWSCVILFVILAGYLPFSETDPLGLYRKINAAPSCPPWFSAEVKFL 240
 DB 181 LSGQGYDGSAAIDWSCVILFVILAGYLPFSETDPLGLYRKINAAPSCPPWFSAEVKFL 240

QY 241 IHRILDENPKTRIOIGIKKDPWFLNXYVPIRAREBEENVLDDTRAVFDGIEGSYVAENV 300
 DB 241 IHRILDENPKTRIOIGIKKDPWFLNXYVPIRAREBEENVLDDTRAVFDGIEGSYVAENV 300

QY 301 ERNDEGLMNAFEMITLSQGLNLSALFDRRQDPVKQTRFVSRRPSEIIANIEAVANS 360
 DB 301 ERNDEGLMNAFEMITLSQGLNLSALFDRRQDPVKQTRFVSRRPSEIIANIEAVANS 360

QY 361 MGFKSHTRNFKTRLEGSSIKAGQLAVIIEYVAPSLFVMDVRKAAGETLEYHKFYKKL 420
 DB 361 MGFKSHTRNFKTRLEGSSIKAGQLAVIIEYVAPSLFVMDVRKAAGETLEYHKFYKKL 420

QY 421 CSKLENIWIWATEGIPKSEILRTITTF 446
 DB 421 CSKLENIWIWATEGIPKSEILRTITTF 446

RESULT 5
 AAG50963
 ID AAG50963 standard; Protein; 442 AA.
 XX
 AC AAG50963;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 64635.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR

PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139452.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.

PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144353.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144354.	PR	14-OCT-1999;	99US-0159638.
PR	20-JUL-1999;	99US-0144632.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160767.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	28-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	26-OCT-1999;	99US-0161362.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	03-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161922.
PR	04-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	05-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149358.			
PR	17-AUG-1999;	99US-0149315.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149920.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151066.			
PR	30-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	29-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			
PR	05-OCT-1999;	99US-0157753.			
Query Match 99.1%; Score 2273; DB 21; Length 442;					
Best Local Similarity 100.0%; Pred. No. 7e-235;					
Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Qy	5	MRRVGKYEVRTIGEGTFAKVFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREISIMK	64		
Db	1	MRRVGKYEVRTIGEGTFAKVFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREISIMK	60		
Qy	65	IVRHPNIVRLYEVLASPSKIYVLEFVTGGELFDRIVHKGRLEESRSRKYFQQLVDVAH	124		
Db	61	IVRHPNIVRLYEVLASPSKIYVLEFVTGGELFDRIVHKGRLEESRSRKYFQQLVDVAH	120		
Qy	125	CHCKGVYHRDLKPENLDDTNGNLKVDSDFGLSALPOBVGVELLRTTCTGPNYVAPEVLSGQ	184		
Db	121	CHCKGVYHRDLKPENLDDTNGNLKVDSDFGLSALPOBVGVELLRTTCTGPNYVAPEVLSGQ	180		
Qy	185	GYDGSADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEYKFLIHR	244		
Db	181	GYDGSADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEYKFLIHR	240		
Qy	245	LDPNKTRIQIQIKKDPWRLNVPIRAREEEVNLDIRAVFDGIEGSYVAENVRND	304		
Db	241	LDPNKTRIQIQIKKDPWRLNVPIRAREEEVNLDIRAVFDGIEGSYVAENVRND	300		
Qy	305	EGELMMNAFEMITLSQGLNLSALFDRRQDFVKQTFVSRREPSEIIANIEAVANGMFK	364		
Db	301	EGELMMNAFEMITLSQGLNLSALFDRRQDFVKQTFVSRREPSEIIANIEAVANGMFK	360		
Qy	365	SHTRNFKTRLEGLSSIKAGQLAVVIEIYVAPSLFMDVVRKAAGETLEYHKFKYKLC	424		
Db	361	SHTRNFKTRLEGLSSIKAGQLAVVIEIYVAPSLFMDVVRKAAGETLEYHKFKYKLC	420		
Qy	425	ENIIWRATEGIPKSEILRTITF	446		
Db	421	ENIIWRATEGIPKSEILRTITF	442		
RESULT 6					
AAG50964					
ID AAG50964 standard; Protein; 405 AA.					

XX AAG50964; 18-JUN-1999; 99US-0139462.
AC 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
DT 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 64636. 23-JUN-1999; 99US-0140353.
XX 24-JUN-1999; 99US-0140354.
KW Protein identification; signal transduction pathway; metabolic pathway; 24-JUN-1999; 99US-0140695.
KW hybridisation assay; genetic mapping; gene expression control; promoter; 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
OS Arabidopsis thaliana. 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 01-JUL-1999; 99US-0142154.
PN EPI033405-A2. 02-JUL-1999; 99US-0142055.
XX 06-JUL-1999; 99US-0142390.
PD 08-JUL-1999; 99US-0142803.
XX 09-JUL-1999; 99US-0142920.
PF 12-JUL-1999; 99US-0142977.
XX 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144633.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.

PR 04-JUN-1999;	99US-0137502.	PR 06-AUG-1999;	99US-0147303.
PR 07-JUN-1999;	99US-0137724.	PR 06-AUG-1999;	99US-0147416.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138847.	PR 10-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 23-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 25-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140695.	PR 07-SEP-1999;	99US-0152363.
PR 24-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140891.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 06-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.
PR 13-JUL-1999;	99US-0143542.	PR 04-OCT-1999;	99US-0157117.
PR 14-JUL-1999;	99US-0143624.	PR 05-OCT-1999;	99US-0157753.
PR 15-JUL-1999;	99US-0144005.	PR 06-OCT-1999;	99US-0157865.
PR 16-JUL-1999;	99US-0144085.	PR 07-OCT-1999;	99US-0158029.
PR 16-JUL-1999;	99US-0144086.	PR 08-OCT-1999;	99US-0158232.
PR 19-JUL-1999;	99US-0144325.	PR 12-OCT-1999;	99US-0158369.
PR 19-JUL-1999;	99US-0144331.	PR 13-OCT-1999;	99US-0158393.
PR 19-JUL-1999;	99US-0144332.	PR 13-OCT-1999;	99US-0159294.
PR 19-JUL-1999;	99US-0144333.	PR 13-OCT-1999;	99US-0159295.
PR 19-JUL-1999;	99US-0144334.	PR 14-OCT-1999;	99US-0159329.
PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144884.	PR 14-OCT-1999;	99US-0159637.
PR 21-JUL-1999;	99US-0144814.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0145086.	PR 18-OCT-1999;	99US-0159584.
PR 22-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160741.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160768.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160770.
PR 23-JUL-1999;	99US-0145145.	PR 21-OCT-1999;	99US-0160814.
PR 23-JUL-1999;	99US-0145216.	PR 22-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160980.
PR 26-JUL-1999;	99US-0145276.	PR 22-OCT-1999;	99US-0160981.
PR 27-JUL-1999;	99US-0145913.	PR 23-OCT-1999;	99US-0160989.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145919.	PR 25-OCT-1999;	99US-0161405.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161406.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161359.
PR 02-AUG-1999;	99US-0146388.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146389.	PR 26-OCT-1999;	99US-0161361.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161320.
PR 04-AUG-1999;	99US-0147204.	PR 28-OCT-1999;	99US-0161992.
PR 04-AUG-1999;	99US-0147302.	PR 28-OCT-1999;	99US-0161993.
PR 05-AUG-1999;	99US-0147192.	PR 29-OCT-1999;	99US-0162142.
PR 05-AUG-1999;	99US-0147260.		

Query Match 64.7%; Score 1484.5; DB 21; Length 445;
Best Local Similarity 62.3%; Pred. No. 4e-150;


```
XX OS Zea mays.
XX XX WO200036115-A2.
XX XX 22-JUN-2000.
XX XX 15-DEC-1999; 99WO-US29824.
XX XX 16-DEC-1998; 98US-0112563.
XX XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX XX Allen SM, Heppard EP, Sakai H, Weng Z, Helentjaris TG, Macool DJ,
XX XX Miao G;
XX XX WPI; 2000-431593/37.
XX XX N-PSDB; AAA52767.
XX XX New nucleic acids encoding sucrose non-fermenting 1 (SNF1) proteins
XX XX involved in carbon catabolite repression in plants and seeds, useful
XX XX for controlling carbon and nitrogen partitioning pathways during plant
XX XX growth and development .
XX XX Claim 10; Page 33-34; 59pp; English.
XX XX
XX CC The present sequence is a putative sucrose non-fermenting protein SNF1
XX CC protein sequence from corn. Its coding sequence was isolated by searching
XX CC a corn endosperm cDNA library for sequences similar to those encoding
XX CC SNF1 in Arabidopsis thaliana, Cucumis sativus, Glycine max and Oryza
XX CC sativa. The protein is involved in carbon catabolite repression, and so
XX CC the gene and protein can be used in plants to control the nitrogen and
XX CC carbon partitioning pathways during plant growth and development. They
XX CC can also be used to alter the accumulation of carbohydrates, lipids and
XX CC proteins during plant growth.
XX XX
XX SQ Sequence 481 AA;

Query Match 64.1%; Score 1469; DB 21; Length 481;
Best Local Similarity 63.2%; Pred. NO. 2.1e-148;
Matches 277; Conservative 74; Mismatches 85; Indels 2; Gaps 15;

Qy 5 MRRVGKVEVGRTGEGTFAKVKFARNTDGDGNVAIKMAKSTILKNRMVDQIKREISIMK 64
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40 LRRVGKVEVGRTGEGTFAKVKFARNTDGTSGVAMKVLDRSSILKNRMVAEQIKREISIMK 99

Qy 65 IVRHPNVLRLVEVLASPSKIYIVLFEFVTGGELFDRIVHKGRLSESESKYFQQLVDVAH 124
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 LVRHPNVRLHEVLASRKKIFILFEFTGGELFDKIIRHGRLSADARRVFQQLIDGVDF 159

Qy 125 CHKGVYHRDLKPNLLDTHGNLUKVSDFGLSALPQBGVVELLRTTCGTPNTVADEVLSGQ 184
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 CHKGVYHRDLKPNLLDTHGNLUKISDFGLSAPQAQGSFLRLTTCGTPNTVADEVLSHK 219

Qy 185 GYDGSADISGCVILFVLILAGLVPFSTDLPGLYRKINAEFSQPPWFSAEVLTHRI 244
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 GYNALADTWSGVLIVLLAGLVPFDEVDUTTLYGKLESSEYFPAWFGAGSLIRRI 279

Qy 245 LDPNPKTRIQGHKKQDPFRNLNYPPIRAREEEVNLDDIRAVFDGIEGYSVARNVERND 304
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
280 LDPNPKTRIRIEIRSDWFQKNVEPIKEIENEENVLDDVNAAFDDPEDD--NEDAFED 337

Qy 305 EGPLMMNAFEMITISQGLNLSALFDRDQDFVYRQTRFVSRREPSEIITANIEAVNSMGFK 364
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
338 TGPLTNADFILISQGLNLAALFDRDQCDKQLQNRFLSRNPAPKAVILSSMEVVAQSMGFK 397

Qy 365 SHTRNFKTRLEGLSSIKAGQAVVIEIYEVAPSLPMVDVRKAAGTLEYHKFKLCSKL 424
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
398 THIRYNKRVGLNADKTSLSLVMVEFEVAPSLFMVELQRAAGDTSEYNTFVNVCGL 457

Qy 425 ENIIWRATEGPKSEILR 442
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
458 DDIWKFTKCKSRIPIR 475
```

```
RESULT 10
ABR40707
ID ABR40707 standard; Protein; 481 AA.
XX
AC ABR40707;
XX
DT 16-MAY-2003 (first entry)
XX
DE Zea mays oil trait related protein sequence SEQ ID NO:234.
XX
KW Plant; oil trait; oil phenotype; altered lipid profile; MAP kinase;
KW receptor-like protein kinase; mitogen activated protein kinase; oil;
KW LIP15-like transcription factor caleosin; ATP citrate lyase; SNF1;
KW CKC-like transcription factor; antisense inhibition; co-suppression;
XX transgenic plant.
XX
OS Zea mays.
XX
PN WO2003002751-A2.
XX
PD 09-JAN-2003.
XX
PF 27-JUN-2002; 2002WO-US20152.
XX
PR 29-JUN-2001; 2001US-301913P.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Allen WB, Cahoon RE, Epelbaum S, Famodu OO, Harvell LT;
PI Jones TJ, Kinney AJ, Klein TM, Li C, Oliveira IC, Sakai H;
PI Shen B, Tarczynski MC;
XX
DR WPI; 2003-201509/19.
DR N-PSDB; ACC00742.
XX
PT Novel nucleotide fragment encoding polypeptides having receptor-like
PT protein kinase activity, caleosin-like activity, useful for altering
PT oil phenotypes in plants such as sunflower, coconut, soybean, wheat and
PT rice .
XX
PS Claim 12; Page 270-272; 542pp; English.
XX
CC The present invention describes an isolated nucleotide fragment (I)
CC comprising a nucleic acid sequence (NS) chosen from a NS encoding a
CC polypeptide (PP) having receptor-like protein kinase activity, mitogen
CC activated protein (MAP)-kinase activity, LIP15-like transcription factor
CC activity, caleosin-like activity, ATP citrate lyase activity, SNF1-like
CC activity and CKC-like transcription factor activity. Also described:
CC (1) complement (II) of (I); (2) a chimeric construct (III) comprising (I)
CC or (II), operably linked to a regulatory sequence; (3) a plant (IV)
CC comprising (III) in its genome; (4) seeds (V) obtained from (IV); and
CC (5) oil obtained from (V); (I) or its part can be used in antisense
CC inhibition or co-suppression in a transformed plant. (III) is useful for
CC altering the oil phenotype in a plant such as corn, soybean, wheat, rice,
CC canola, Brassica, sorghum, sunflower or coconut. (III) is also useful for
CC creating transgenic plants having altered lipid profiles. (I) can also
CC be used as a hybridisation probe. ACC00626 to ACC00868 and ABR40591 to
CC ABR40879 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 481 AA;

Query Match 64.1%; Score 1469; DB 24; Length 481;
Best Local Similarity 63.2%; Pred. No. 2.1e-148;
Matches 277; Conservative 74; Mismatches 85; Indels 2; Gaps 1;

Qy 5 MRRVGKVEVGRTGEGTFAKVKFARNTDGDGNVAIKMAKSTILKNRMVDQIKREISIMK 64
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40 LRRVGKVEVGRTGEGTFAKVKFARNTGTSGVAMKVLDRSSILKNRMVAEQIKREISIMK 99
```


PR	18-JUN-1999;	99US-0139452.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142085.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	26-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	28-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145199.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147192.	Query Match 58.6%; Score 1344.5; DB 21; Length 416;		
PR	05-AUG-1999;	99US-0147260.	Best Local Similarity 64.8%; Pred.No. 4e-135;		
PR	06-AUG-1999;	99US-0147303.	Matches 250; Conservative 69; Mismatches 66; Indels 1; Gaps 1;		
PR	06-AUG-1999;	99US-0147316.	QY	5	MREVKYEVGRITGEGTFAKVKFARNTDTCNVAIKMAKSTILKKNMVDQIKRETSIMK 64
PR	08-AUG-1999;	99US-0147493.	Db	3	VRKVGVELGRITGEGTFAKVKFAQNTETGESVAMKIVDRSTIILKKNVDQIKRETSIMK 62
PR	09-AUG-1999;	99US-0147935.	QY	65	IVRHPNIVLYEVLASPSKIYIVLEPVTGGELFDRIVHKGRLSEESRKYFOOLVDAVAH 124
PR	10-AUG-1999;	99US-0148171.	Db	63	LVRHPCVRLYEVLASRTKIYILEYITGGELFDKIVRNGRLSESEARKYFHQIDVDY 122
PR	11-AUG-1999;	99US-0148319.	QY	125	CHCKGVYHRDLKPNLLDTNGNLKYSDFGLSALPOEGVELLETTCCGTNNYVAPYLSGQ 184
PR	12-AUG-1999;	99US-0148341.	Db	123	CHSKGVYHRDLKPNLLDQGNLKISDFGLSALPEQGVITILKTCCGTNNYVAPYLSHK 182
PR	13-AUG-1999;	99US-0148565.	QY	185	GYDGSADIWSCGVILFVILAGYLPFSETDLPGLYRKINAARFSCPPWFSAEVKFLIHR 244
PR	13-AUG-1999;	99US-0148684.	Db	183	GYNGAVADIWSCGVILYVLWAGYLPFDEMDLPTLYSKIDKAEFSCPSYFALGAKSLINRI 242
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			

```
Oy 245 LDPNPKTRIQIGIKDPWFLNVPTRAREEEVNLDDIRAVFDGIEGYSVAENVERND 304
Db 243 LDPNPKTRITIAERKDEWFLKDYTPQLIDYEHVNLDDVYAADFDPPEQTYAQDGR-D 301
Oy 305 EGPLMMNAFEMITISQGLNLSALFDRDQFVKROTFRVSRPEPFIITANIEAVANSNGFK 364
Db 302 TGPTILNAFDLIIISQGLNLATLFDRGKSKHQIRFLSHKPAENVLSSMEVVSQSMGFK 361
Oy 365 SHTNFKTRLEGLSIXKAGQAVVIE 390
Db 362 THIRNYKRVBGLSANKTSHPSVILE 387

RESULT 14
AAG31349
ID AAG31349 standard; Protein; 409 AA.
XX
AC AAG31349;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37633.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
```

Query Match 58.3%; Score 1336.5; DB 21; Length 409;
Best Local Similarity 60.9%; Pred. No. 2.8e-134;

The present invention describes an isolated nucleotide fragment (I) comprising a nucleic acid sequence (NS) chosen from a NS encoding a polypeptide (PP) having receptor-like protein kinase activity, mitogen activated protein (MAP)-kinase activity, LIP15-like transcription factor activity, caldesin-like activity, ATP citrate lyase activity, SNF1-like activity and CKC-like transcription factor activity. Also described: (1) complement (II) of (I); (2) a chimeric construct (III) comprising (I) or (II), operably linked to a regulatory sequence; (3) a plant (IV) comprising (III) in its genome; (4) seeds (V) obtained from (IV); and (5) oil obtained from (V). (I) or its part can be used in antisense inhibition or co-suppression in a transformed plant. (III) is useful for altering the oil phenotype in a plant such as corn, soybean, wheat, rice, canola, brassica, sorghum, sunflower or coconut. (III) is also useful for creating transgenic plants having altered lipid profiles. (I) can also be used as a hybridisation probe. ACC00626 to ACC00868 and ABR40591 to ABR40879 represent sequences used in the exemplification of the present invention.

AA	Sequence	441 AA;
SQ		

Query Match 57.3%; Score 1314; DB 24; Length 441;
Best Local Similarity 57.7%; Pred. No. 8.4e-132;
Matches 248; Conservative 81; Mismatches 95; Indels 6; Gaps 3;

QY 6 RRVGKYEVGRTIGEGTFAKVVFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREISIMKI 65

Db 9 RRVGKYEVRTIGEGTFAKVVFARNSETGEPVALKILDKEKVLKHKMAEQIRREIATMKL 68

QY 66 VRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESESRKYFQQLVDAVAHC 125

Db 69 IKHPNVVQLYEVMSAKTKIFIIILEYVTGGELFDKI VNDGRMKEDARRFYQQLIHAVDYC 128

QV 126 HCKGVYHRDLKPENLLLDTNGNLKVSD⁶FLSALPOEGVE--LLRTTCGTPNYVAPEVLSG 183

db
129 HSRGVYHRDLKPENLLD SYGNLKI SDFGLSALSOOVRDDGLLHTSCGTPNYVAPVLLND 188

Ov 184 OGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHR 243

Db 189 RGYD GATAD MWSCGVLYVLLAGYLPEDDSNLMNLYKISSGEFNCPPWLSIGAMKLTTR 248

QV 244 II:DPNPKTRITOTOGTKKDPWFRINVVPTAREREREVNI.DDTRAVFDGTGEGSYVAENVERN 303

db 249 II.DPNMTRVTPOEVEFEDEWEKKDYKPPVFEERDDSNMDDTDAVFKDSEEHIV---TEKR 305

QV 304 DEGPLMNAFEMITLSOGLNLSALEFDRRODFVKROTRFVSRREPSEIIANIEAVANSMGF 363

Db 306 EEOPAAINAFETISMSRGLNLENLEDPDEOFF-KRETRITLRGGANETIEKIEEAAKPLGF 364

OV 364 KSHTRNEKTRIEGLSSIKAGOLAVVTEIYEVAPSLEMVDPVKAAGETLEYHKFYKKLCSK 423

365 DVOKNYKMRLENVKA GRKGNLNVATEI FOVAPSLHMVOVSKSGDTEFFHKEYKLLSNS 424

424 LENI IWRATE 433 Ov

Db 425 LEOVWWTNNE 434

Search completed: October 20, 2003, 16:50:46
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2003, 01:18:57 ; Search time 18159 Seconds
(without alignments)
11588.695 Million cell updates/sec

Title: US-09-824-735-1

Perfect score: 5144

Sequence: 1 tttttgcgaggaaatctcg.....acgttttgatcccaacttaa 5144

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.em.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rtd.*

36: em.htg.man.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	5144	100.0	5144	8	AF237670	AF237670 Arabidops
2	5144	100.0	50015	8	AB025611	AB025611 Arabidops
3	4527	88.0	16228	8	F6I13	AF262044 Arabidops
C 4	459.4	8.9	110469	8	AC006258	AC006258 Arabidops
C 5	459.4	8.9	110680	8	AC006259	AC006259 Arabidops
C 6	457.6	8.9	105223	8	AC007399	AC007399 Arabidops
C 7	457.6	8.9	157000	8	AC007627	AC007627 Genomic S
C 8	456.2	8.9	133195	8	AC007123	AC007123 Arabidops
C 9	455.2	8.8	87581	8	ATT7H20	AL162508 Arabidops
C 10	453.4	8.8	158096	8	AC007887	AC007887 Genomic s
C 11	451.8	8.8	108767	8	AC011809	AC011809 Arabidops
C 12	449	8.7	80139	8	AC027135	AC027135 Arabidops
C 13	449	8.7	107200	8	AC006551	AC006551 Arabidops
C 14	449	8.7	110664	8	AC074360	AC074360 Arabidops
C 15	448	8.7	88643	8	AC022522	AC022522 Arabidops
C 16	448	8.7	123386	8	F12F1	AC002131 Arabidops
C 17	446.2	8.7	82652	8	AC007020	AC007020 Arabidops
C 18	445.4	8.7	86139	8	AB028616	AB028616 Arabidops
C 19	442.4	8.6	126386	8	AC005169	AC005169 Arabidops
C 20	441.8	8.6	82451	8	AB073160	AB073160 Arabidops
C 21	439.8	8.5	35308	8	AC079285	AC079285 Arabidops
C 22	439.8	8.5	63773	8	AC074025	AC074025 Arabidops
C 23	436.4	8.5	82053	8	AC022455	AC022455 Arabidops
C 24	435.8	8.5	95581	8	AC015448	AC015448 Arabidops
C 25	433.2	8.4	111767	8	F21E10	AF058914 Arabidops
C 26	430.4	8.4	77636	8	AB077822	AB077822 Arabidops
C 27	430.4	8.4	86014	8	F19C14	AC008051 Sequence
C 28	430.2	8.4	99053	8	AC005292	AC005292 Genomic s
C 29	430	8.4	108387	8	AC016662	AC016662 Arabidops
C 30	429.4	8.3	81580	8	AP000736	AP000736 Arabidops
C 31	429.4	8.3	98017	8	AC027033	AC027033 Arabidops
C 32	429.4	8.3	98948	8	ATF6L21	AC096882 Arabidops
C 33	429.4	8.3	198301	8	ATCHRIV31	AL161531 Arabidops
C 34	428.4	8.3	90284	8	T2K10	AC005966 Arabidops
C 35	427.6	8.3	140680	8	ATF13112	AL132292 Arabidops
C 36	427.2	8.3	84710	8	AB026643	AB026643 Arabidops
C 37	426.6	8.3	101933	8	AC004146	AC004146 Arabidops
C 38	425.6	8.3	96899	8	AC013288	AC013288 Arabidops
C 39	423.6	8.2	118737	8	T3H13	AF128396 Arabidops
C 40	423.6	8.2	179771	8	ATCHRIV25	AL161513 Arabidops
C 41	422.2	8.2	68889	8	AB010694	AB010694 Arabidops
C 42	421.8	8.2	57246	8	AB010069	AB010069 Arabidops
C 43	419.2	8.1	118737	8	T3H13	AF128396 Arabidops
C 44	419.2	8.1	179771	8	ATCHRIV25	AL161513 Arabidops
C 45	419	8.1	93443	8	ATF2206	AL050300 Arabidops

ALIGNMENTS

RESULT 1
AF237670
LOCUS
DEFINITION
Arabidopsis thaliana serine/threonine protein kinase SOS2 (SOS2)
ACCESSION
AF237670
VERSION
AF237670.1 GI:7453644
KEYWORDS
Arabidopsis thaliana (thale cress)
SOURCE
Arabidopsis thaliana
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 5144)
AUTHORS
Liu,J., Ishitani,M., Halfter,U., Kim,C.S. and Zhu,J.K.

Pred. No. is the number of results predicted by chance to have a

TITLE The Arabidopsis thaliana SOS2 gene encodes a protein kinase that is required for salt tolerance
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3730-3734 (2000)
MEDLINE 20202704
PUBMED 10725382
2 (bases 1 to 5144)
AUTHORS Liu, J., Ishitani, M. and Zhu, J.-K.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2000) Plant Sciences, University of Arizona, Forbes 303, Tucson, AZ 85721, USA
FEATURES
source Location/Qualifiers
1. 5144
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/culturvar="Columbia"
/db_xref="taxon:3702"
/chromosome="v"
/map="between 26D22-1 and LEMOK9 on BAC K21B8"
<2004. .>5133
/gene="SOS2"
/note="salt overly sensitive 2"
join(2004. .2168,2255. .2317,2416. .2487,2573. .2680, 2778. .2852,2921. .3100,3225. .3314,3418. .3540,4052. .4171, 4400. .4516,4735. .4791,4879. .4953,5038. .>5133)
/gene="SOS2"
/product="serine/threonine protein kinase SOS2"
join(2004. .2168,2255. .2317,2416. .2487,2573. .2680, 2778. .2852,2921. .3100,3225. .3314,3418. .3540,4052. .4171, 4400. .4516,4735. .4791,4879. .4953,5038. .5133)
/gene="SOS2"
/note="forms a kinase complex with calcium sensor SOS3"
/codon_start=1
/product="serine/threonine protein kinase SOS2"
/protein_id="AA62923.1"
/db_xref="GI:7453645"

translation="MTKMRVRGVKYEVRTIGRTFAKVKFARNTDTGNVALKINAK
STLXRMVDQIKRETSIMKIVRPNIIVLYELVAFSPKIIYVLEFVTGELFDRIVH
KGLBESERKVFQOVLDAVAHCKGVVHRDLKPNLLDNTGNLKVSDPGLSALPO
EGVELLRTCTGVNVVAPVLSGGYDGSNAIDWSGVILFVILAGVLPSEDLFGL
YRKINAEKSCFWFAEVKPLIHLRIDNPVKRIQIQGKIDPWFRLNYPVIRABE
EVLNLDIRAFVDFIGSVYVAENVRNDEGLMNAFEMITLSQGLNLSALFDRQDF
VKQTFVRRESEIIANIEAVANSFGKSHTRNPKRLRLEGLSSIKAGQLAVIEIY
EVAPSLFMDVDRKAAGETLHYKFKLCKLENIIWRATEGIPKSEILRTITF"
BASE COUNT 1523 a 851 c 943 g 1827 t
ORIGIN

Query Match 100.0%; Score 5144; DB 8; Length 5144;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTTTGGCGGAAAATCTCGGTTTACGTTTTTGGCGGGAATACTCGGTTTACGTTT 60
DB 1 TTTTGGCGGAAAATCTCGGTTTACGTTTTTGGCGGGAATACTCGGTTTACGTTT 60
QY 61 TGGCGGAAAATCTCGGTTTACGTTTTTGGCGGAAAATCAACGGTTTACGTTT 120
DB 61 TGGCGGAAAATCTCGGTTTACGTTTTTGGCGGAAAATCAACGGTTTACGTTT 120
QY 121 CGGGAATAATCAACGGTTTACGTTTTTGGGAAAATCAACGGTTTACGTTTCAAT 180
DB 121 CGGGAATAATCAACGGTTTACGTTTTTGGGAAAATCAACGGTTTACGTTTCAAT 180
QY 181 TTCATTGCTTGATATTAAAGAAATTTGGAAAATTAATTAATTAATTAATTAATTA 240
DB 181 TTCATTGCTTGATATTAAAGAAATTTGGAAAATTAATTAATTAATTAATTAATTA 240
QY 241 TGTGTTGGTTAAACCTAAATTTGGCAATTTAGAGATTTAGTGGTTTATTCATTT 300
DB 241 TGTGTTGGTTAAACCTAAATTTGGCAATTTAGAGATTTAGTGGTTTATTCATTT 300
QY 301 TACAAAATTTGATGGTTAAATGGATAAACCAATGGAACCAATTAACCACTAAC 360
DB 301 TACAAAATTTGATGGTTAAATGGATAAACCAATGGAACCAATTAACCACTAAC 360

QY 361 TCATTTTACTCATCAACCAATTTGACTCATCACTCACTCACTCACTCACTCACTCACT 420
DB 361 TCATTTTACTCATCAACCAATTTGACTCATCACTCACTCACTCACTCACTCACTCACT 420
QY 421 TCAAAATTTCACTCATTTAGGGTTCAAGGGTTGAGTTGAGTTGAGTTGAGTTGAGTT 480
DB 421 TCAAAATTTCACTCATTTAGGGTTCAAGGGTTGAGTTGAGTTGAGTTGAGTTGAGTT 480
QY 481 TTGACCCATTTTGACACCCCTACATATGATCATCAAGTTAATAATCAAAAATTTACTATTGA 540
DB 481 TTGACCCATTTTGACACCCCTACATATGATCATCAAGTTAATAATCAAAAATTTACTATTGA 540
QY 541 TAATCTTTTAAACCGAATTTGTTTGTAAAGTTTCATTTACGTTATTTATATATAAAAAACATA 600
DB 541 TAATCTTTTAAACCGAATTTGTTTGTAAAGTTTCATTTACGTTATTTATATATAAAAAACATA 600
QY 601 ATGCAAAAGTAAATGATATAGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 660
DB 601 ATGCAAAAGTAAATGATATAGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 660
QY 661 TACTTTTTTCACTTCTTATCAATATTTTCTTATTTTCTTATTTTCTTATTTCTTATTTCTTAT 720
DB 661 TACTTTTTTCACTTCTTATCAATATTTTCTTATTTTCTTATTTTCTTATTTCTTATTTCTTAT 720
QY 721 TCATATTTCTTTCATCAATTCATTTTCTTGTCTTGTATTTCTTGTATTTCTTGTATTTCTTGTAT 780
DB 721 TCATATTTCTTTCATCAATTCATTTTCTTGTCTTGTATTTCTTGTATTTCTTGTATTTCTTGTAT 780
QY 781 TGCCTTTTCTTTCATTTGACATTTGTCATCGTTTGTATTTGTATTTGTATTTGTATTTGTATTT 840
DB 781 TGCCTTTTCTTTCATTTGACATTTGTCATCGTTTGTATTTGTATTTGTATTTGTATTTGTATTT 840
QY 841 TGTGAATCTCCACACCGACTATAGGACTATCAAGAAGTTTTCATAAAGTTTTCATAAAGTTTTC 900
DB 841 TGTGAATCTCCACACCGACTATAGGACTATCAAGAAGTTTTCATAAAGTTTTCATAAAGTTTTC 900
QY 901 TGAACCTTTAAAGCAAAATCCCAACCGACTATAGAACCATAGTCTCTTCTTCTTCTTCTTCTTCT 960
DB 901 TGAACCTTTAAAGCAAAATCCCAACCGACTATAGAACCATAGTCTCTTCTTCTTCTTCTTCTTCT 960
QY 961 AGCTCTTCTCGTCCGCGGCAAAATCTAGGCCATAAAGCTCTTCAACATCACTAGTAT 1020
DB 961 AGCTCTTCTCGTCCGCGGCAAAATCTAGGCCATAAAGCTCTTCAACATCACTAGTAT 1020
QY 1021 ATTGACCGTGAGCATCTTTTGTGACCATTTGCTTTGTGAATGAACCGTCAATAAAGCTGTT 1080
DB 1021 ATTGACCGTGAGCATCTTTTGTGACCATTTGCTTTGTGAATGAACCGTCAATAAAGCTGTT 1080
QY 1081 ATCACTTACGCAAAATTTTCCCTAGTGTGTTTCCAAACTCTCAACATCTCTTATTCGA 1140
DB 1081 ATCACTTACGCAAAATTTTCCCTAGTGTGTTTCCAAACTCTCAACATCTCTTATTCGA 1140
QY 1141 ACTTTTTTATATATCACTTTGTAGCATTTGGAAGTATCTTTGTATGCTTTGCTTTTAAACT 1200
DB 1141 ACTTTTTTATATATCACTTTGTAGCATTTGGAAGTATCTTTGTATGCTTTGCTTTTAAACT 1200
QY 1201 TAGACATCTCTTCTTCTTGTGTTTGTGAAACCTTTGCTTGAATGAAGTTTAAAGTT 1260
DB 1201 TAGACATCTCTTCTTCTTGTGTTTGTGAAACCTTTGCTTGAATGAAGTTTAAAGTT 1260
QY 1261 TTGTAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTT 1320
DB 1261 TTGTAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTT 1320
QY 1321 TTGAAAATTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTT 1380
DB 1321 TTGAAAATTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTT 1380
QY 1381 CTAAATTTGATAGTATTTAGATTTAGATTTAGATTTAGATTTAGATTTAGATTTAGATTT 1440
DB 1381 CTAAATTTGATAGTATTTAGATTTAGATTTAGATTTAGATTTAGATTTAGATTTAGATTT 1440

Qy	1441	ACAA	TTATTAATTAATGAAAGATTCAAATGCAATGTTAGTTTAAAAAGTTTAATCACCCA	1500
Db	1441	ACAA	TTATTAATTAATGAAAGATTCAAATGCAATGTTAGTTTAAAAAGTTTAATCACCCA	1500
Qy	1501	AACGG	ATAATTGACCCGAACGACTAAATTCAGGTGCTATACGGGTACAGTCAAATAACCCG	1560
Db	1501	AA	CGGATAATTGACCCGAACGACTAAATTCAGGTGCTATACGGGTACAGTCAAATAACCCG	1560
Qy	1561	ACCG	AAATGCTCAACGGATCTGCACGAGCGTTCCGATCGACGATTCACGGCTTTCGCGA	1620
Db	1561	ACCG	AAATGCTCAACGGATCTGCACGAGCGTTCCGATCGACGATTCACGGCTTTCGCGA	1620
Qy	1621	ATGCG	ATCAGAGCGCTTCCTTCTCTCTAAGCTTTCATCAACCCCTTCTCTCGAATC	1680
Db	1621	ATGCG	ATCAGAGCGCTTCCTTCTCTCTAAGCTTTCATCAACCCCTTCTCTCGAATC	1680
Qy	1681	CAAT	TCGGTATTTACAGATTCCTTTTCGACTAGCCCAATCGCAATTCAGCCGTACGA	1740
Db	1681	CAAT	TCGGTATTTACAGATTCCTTTTCGACTAGCCCAATCGCAATTCAGCCGTACGA	1740
Qy	1741	ATT	TTATATATGATTAATGATTCCTTTCGACTAGCCCAATCGCAATTCAGCCGTACGA	1800
Db	1741	ATT	TTATATATGATTAATGATTCCTTTCGACTAGCCCAATCGCAATTCAGCCGTACGA	1800
Qy	1801	ATG	TTGATTTATGCTGCGAAATCAAATTTATGAGTAAATTTGATATATGATG	1860
Db	1801	ATG	TTGATTTATGCTGCGAAATCAAATTTATGAGTAAATTTGATATATGATG	1860
Qy	1861	GTG	TAGGAGCTAAATTTAGCAAAACAATACCTGTAAGTAAATTTGAAATCGAAATCA	1920
Db	1861	GTG	TAGGAGCTAAATTTAGCAAAACAATACCTGTAAGTAAATTTGAAATCGAAATCA	1920
Qy	1921	TTA	AGGAAAGAGGTTTTTACTAAGTCTCTGAAATCTGATGATAGTTTGTGTCATG	1980
Db	1921	TTA	AGGAAAGAGGTTTTTACTAAGTCTCTGAAATCTGATGATAGTTTGTGTCATG	1980
Qy	1981	GAT	CAGATAAAAGTTTGTAAAGAAATGCAAGAAATGAGAGTGGGCAAGTACGAGG	2040
Db	1981	GAT	CAGATAAAAGTTTGTAAAGAAATGCAAGAAATGAGAGTGGGCAAGTACGAGG	2040
Qy	2041	TTG	GTGCAAAATAGTGTAGGAGCACTTTGCTAAGGTTAAGTTTGCAGGAAACAGACA	2100
Db	2041	TTG	GTGCAAAATAGTGTAGGAGCACTTTGCTAAGGTTAAGTTTGCAGGAAACAGACA	2100
Qy	2101	CTG	GTGATTAATGTAGCCATCAAAATTAATGCTTAAGGTACAATTAAGCAACAGAAATG	2160
Db	2101	CTG	GTGATTAATGTAGCCATCAAAATTAATGCTTAAGGTACAATTAAGCAACAGAAATG	2160
Qy	2161	TTG	ATCAGGTATGTTCTGGAATGTTTTTATCAGGAAATTAAGGTTGTGCGTCAATG	2220
Db	2161	TTG	ATCAGGTATGTTCTGGAATGTTTTTATCAGGAAATTAAGGTTGTGCGTCAATG	2220
Qy	2221	ATG	ATCTTTGATTTGTTTTAAAGCTCTTTTACAGATAAAAGAGAGATCTATAATGAA	2280
Db	2221	ATG	ATCTTTGATTTGTTTTAAAGCTCTTTTACAGATAAAAGAGAGATCTATAATGAA	2280
Qy	2281	GAT	TGTTGCTCACCCGAAACATAGTGAAGTTGATGAGGTATGTTGTTGTTTCCATGCA	2340
Db	2281	GAT	TGTTGTTGCTCACCCGAAACATAGTGAAGTTGATGAGGTATGTTGTTGTTTCCATGCA	2340
Qy	2341	TC	TGCGAAATTTATCTCTGAAGTGTGTTTGGCATCATGTTCTCTGTTGTTTTTTG	2400
Db	2341	TC	TGCGAAATTTATCTCTGAAGTGTGTTTGGCATCATGTTCTCTGTTGTTTTTTG	2400
Qy	2401	AT	TTTTCCGATGTAGTGTGGGAGTCCTTCGAAATATATATAGTTTGGAGTTTGTG	2460
Db	2401	AT	TTTTCCGATGTAGTGTGGGAGTCCTTCGAAATATATATAGTTTGGAGTTTGTG	2460
Qy	2461	ACA	GAGGAGAGCTCTTTTGATAGAAATGTACGGAACTTCCATATCTTGTAGGACGCTCA	2520
Db	2461	ACH	GAGGAGAGCTCTTTTGATAGAAATGTACGGAACTTCCATATCTTGTAGGACGCTCA	2520
Qy	2521	TT	AGTTAAACCTCTACTTAATTTTTTAATATATGAAATCTTTCAACAGGTTTCATAA	2580

2521	TTAGTTTAAAAACCTCTCTACTTAAATTTTTTAATATAGAAATCTTTCATGCAGGTTTCATAA	2580
2581	AGGAGGCTTGAAGAAAGTAGTCTCGGAAATATCTTTCACAGCTGTAGATGCTGTGTGC	2640
2581	AGGAGGCTTGAAGAAAGTAGTCTCGGAAATATCTTTCACAGCTGTAGATGCTGTGTGC	2640
2641	TCATTCTCATCTCAAGGGTGTTTACCACCGTGACCTTAAAGGTAAGAGCGTGTTTTGTGTT	2700
2641	TCATTCTCATCTCAAGGGTGTTTACCACCGTGACCTTAAAGGTAAGAGCGTGTTTTGTGTT	2700
2701	ACCAATATTCTTCAGAAATATCTCATCTCGTGTGCAATCCAGACTTCATATTTTGTGTGCG	2760
2701	ACCAATATTCTTCAGAAATATCTCATCTCGTGTGCAATCCAGACTTCATATTTTGTGTGCG	2760
2761	TATGTTATGTTATCTAGCCAGAAATCTTTTACTCGATACAAATGCAAAATCTGAAGGTTT	2820
2761	TATGTTATGTTATCTAGCCAGAAATCTTTTACTCGATACAAATGCAAAATCTGAAGGTTT	2820
2821	CGGATTTCCGACTCAGTGCATGCTCCAGGAAGTAGTCTCTATCTCTCGCTTCAGCAG	2880
2821	CGGATTTCCGACTCAGTGCATGCTCCAGGAAGTAGTCTCTATCTCTCGCTTCAGCAG	2880
2881	CTGCTTACGTGGTCAATTAATCTTATATATCTCAATCAGGAGTAGAACTTCTCGGTAC	2940
2881	CTGCTTACGTGGTCAATTAATCTTATATATCTCAATCAGGAGTAGAACTTCTCGGTAC	2940
2941	CACATGTGGAACCTCGAACTATGTAAGTCCAGAGTATTAAGTGACAGGGTTACGATGG	3000
2941	CACATGTGGAACCTCGAACTATGTAAGTCCAGAGTATTAAGTGACAGGGTTACGATGG	3000
3001	TTCCAGCAGCTGATATTTGTCCTCGGGGTATTTCTTTTCGTTATATTGGCTGGAATATTT	3060
3001	TTCCAGCAGCTGATATTTGTCCTCGGGGTATTTCTTTTCGTTATATTGGCTGGAATATTT	3060
3061	ACCTTTTCCGAGACGGATCTTCCAGGGTTGTACAGAAAAGTAAGTAACATATCTTTGG	3120
3061	ACCTTTTCCGAGACGGATCTTCCAGGGTTGTACAGAAAAGTAAGTAACATATCTTTGG	3120
3121	GAAGAAATCATGAATTCCTTGTGTATGCTTTGTGTAACCGTTTATGATTTGGTTTTGCG	3180
3121	GAAGAAATCATGAATTCCTTGTGTATGCTTTGTGTAACCGTTTATGATTTGGTTTTGCG	3180
3181	AAATTTCCACCTTAGTTTGGCTTTTACACATGTTATTTACAGATAAATCCAGCAGAT	3240
3181	AAATTTCCACCTTAGTTTGGCTTTTACACATGTTATTTACAGATAAATCCAGCAGAT	3240
3241	TTTCTGTCCACCGTGGTTTTCCGCGAGAAGTGAAGTTTTTAATACATAGGATATCTGACC	3300
3241	TTTCTGTCCACCGTGGTTTTCCGCGAGAAGTGAAGTTTTTAATACATAGGATATCTGACC	3300
3301	CCAAATCCCAAACACGTCAGTATTTTGGTTTGTCTCTCTAGCTATCAGGTTTTGGTGAT	3360
3301	CCAAATCCCAAACACGTCAGTATTTTGGTTTGTCTCTCTAGCTATCAGGTTTTGGTGAT	3360
3361	ATTTAAATGTTCTAGTAATATATCTGTTTATCTATTTGTTTCTCAATTAGACACCGT	3420
3361	ATTTAAATGTTCTAGTAATATATCTGTTTATCTATTTGTTTCTCAATTAGACACCGT	3420
3421	ATTCAAATTCAGGAATCAAGAAAGATCCCTGGTTCCAGATTAATATGTCCTTACGTA	3480
3421	ATTCAAATTCAGGAATCAAGAAAGATCCCTGGTTCCAGATTAATATGTCCTTACGTA	3480
3481	GCAAGGGAAGAAAGAAAGTGAATTTGGATGATATTCGTGCAGTTTTTGTGAATTGAG	3540
3481	GCAAGGGAAGAAAGAAAGTGAATTTGGATGATATTCGTGCAGTTTTTGTGAATTGAG	3540
3541	GTTTGTGCTTCGCCCTTCATTTATATGCTCTTTCTCGTCAATTCATTTAAATGTTAAGAT	3600
3541	GTTTGTGCTTCGCCCTTCATTTATATGCTCTTTCTCGTCAATTCATTTAAATGTTAAGAT	3600
3601	CTTCTTAGGAACGTTTGATGACCAAGAAAGTGTTCCTACAGGATAGACAAATAGT	3660

3601 CTCTTAGGAAGCTTTGGATGACCAAGAGAGTCTTTGCTACAGGATAGAACAAAATAGT 3660
 3661 AGGATGTTAGTATCCAAACCTGTAAACCTGCTCTTTATCTCAATTCGCCAACCATAG 3720
 3661 AGGATGTTAGTATCCAAACCTGTAAACCTGCTCTTTATCTCAATTCGCCAACCATAG 3720
 3721 ACCTTAGGAAGACTTAGATTTTACAGAGAAATCTCTATCTCGACCAAAAACCTAGACA 3780
 3721 ACCTTAGGAAGACTTAGATTTTACAGAGAAATCTCTATCTCGACCAAAAACCTAGACA 3780
 3781 AAATCCAGAAATACCCCTAGGCTAATTAACAATGTTCAATGTTACCTATCAATATATATCTCC 3840
 3781 AAATCCAGAAATACCCCTAGGCTAATTAACAATGTTCAATGTTACCTATCAATATATATCTCC 3840
 3841 TGCTAAGATTTGCTTTAGCTTTGGTATAGCTTAGATATATATATATATATATATATATAT 3900
 3841 TGCTAAGATTTGCTTTAGCTTTGGTATAGCTTAGATATATATATATATATATATATATAT 3900
 3901 GCATATTTAGAAATGTTTACTTAACTCAGGCTCTGCTCTTCTTAAACCTTGTTACTTCAA 3960
 3901 GCATATTTAGAAATGTTTACTTAACTCAGGCTCTGCTCTTCTTAAACCTTGTTACTTCAA 3960
 3961 TTGTTTAACTTAAACCTCAGTATCTGCTTAGCTTAAAGTACTTCTTACTGTTTCTTCAAT 4020
 3961 TTGTTTAACTTAAACCTCAGTATCTGCTTAGCTTAAAGTACTTCTTACTGTTTCTTCAAT 4020
 4021 AAGTTGACCTGTCAATTTGTTTCAAGGCGAGTATATAGCGGAGATGTTAGAGAG 4080
 4021 AAGTTGACCTGTCAATTTGTTTCAAGGCGAGTATATAGCGGAGATGTTAGAGAG 4080
 4081 AAATGATGAAGGCGCCCTGATGATGAATGCTTTGAGATGATTAACCTTATCACAGGCTT 4140
 4081 AAATGATGAAGGCGCCCTGATGATGAATGCTTTGAGATGATTAACCTTATCACAGGCTT 4140
 4141 AAATTTATCTGCACTATTTGACAGCGCAGAGTAGTACCTGATTTCTTATTTACTGGTCA 4200
 4141 AAATTTATCTGCACTATTTGACAGCGCAGAGTAGTACCTGATTTCTTATTTACTGGTCA 4200
 4201 AGAGATCTCCATTTGCAATTAAGAAATGTCGGTAGCATCTATTCTTCACTGCTGCTT 4260
 4201 AGAGATCTCCATTTGCAATTAAGAAATGTCGGTAGCATCTATTCTTCACTGCTGCTT 4260
 4261 TGACTGCTTTATGATGCTGCTGCTTCTTATTAATACTAATTAAGTCTTATTTAGATGA 4320
 4261 TGACTGCTTTATGATGCTGCTGCTTCTTATTAATACTAATTAAGTCTTATTTAGATGA 4320
 4321 TTGGTTGCTATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4380
 4321 TTGGTTGCTATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4380
 4381 TGATCTTTGATCTCGAGATTTGTTTAAAGGCAAAACCCGTTTGTGTTTCTCGAAGGGA 4440
 4381 TGATCTTTGATCTCGAGATTTGTTTAAAGGCAAAACCCGTTTGTGTTTCTCGAAGGGA 4440
 4441 ACCTAGTAGAATAATGCTTAACTAGGCTGCTAGCAACTCAATGGGTTTAAAGTCTCA 4500
 4441 ACCTAGTAGAATAATGCTTAACTAGGCTGCTAGCAACTCAATGGGTTTAAAGTCTCA 4500
 4501 TACAGAAACTTCAAGTAAAGATCTTACATTTACATTTACATTTACATTTACATTTACAT 4560
 4501 TACAGAAACTTCAAGTAAAGATCTTACATTTACATTTACATTTACATTTACATTTACAT 4560
 4561 TTTATTTAAACCTCTCAACTGTTTAAACCGATGTTAGATGATGATGATGATGATGATGATG 4620
 4561 TTTATTTAAACCTCTCAACTGTTTAAACCGATGTTAGATGATGATGATGATGATGATGATG 4620
 4621 AAATCTAGTTATATCTTGGGTTTCAAAAAGTCTTAAAGTCTTAAAGTCTTAAAGTCTT 4680
 4621 AAATCTAGTTATATCTTGGGTTTCAAAAAGTCTTAAAGTCTTAAAGTCTTAAAGTCTT 4680
 4681 AATGTTAGCAGAGTACATTTGTTATGATTTGTTACGCTATGCTGCTTCAACAGCAAGG 4740
 4681 AATGTTAGCAGAGTACATTTGTTATGATTTGTTACGCTATGCTGCTTCAACAGCAAGG 4740

4741 CTCGAGGATTTATCTTCGATCAAGGCGGACAGTTAGCTGTGTGTATAGAGTAATTATT 4800
 4741 CTCGAGGATTTATCTTCGATCAAGGCGGACAGTTAGCTGTGTGTATAGAGTAATTATT 4800
 4801 GCTTGTGTGATTTGATTAATTAAGTTTGTCTTCTGCTTCAAGTCTTAAAGGTTCTAGCAAAATTG 4860
 4801 GCTTGTGTGATTTGATTAATTAAGTTTGTCTTCTGCTTCAAGTCTTAAAGGTTCTAGCAAAATTG 4860
 4861 AAATTAACCTACATCAAGATTTTACGAGTGGCACCATCGCTTTTTCATGCTAGAGCTAAGA 4920
 4861 AAATTAACCTACATCAAGATTTTACGAGTGGCACCATCGCTTTTTCATGCTAGAGCTAAGA 4920
 4921 AAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4980
 4921 AAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4980
 4981 AGTTCGATCAATTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5040
 4981 AGTTCGATCAATTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5040
 5041 TACAAGAACTATGTTTGAACCTGGAACCTGGAACCTGGAACCTGGAACCTGGAACCTGGAAC 5100
 5041 TACAAGAACTATGTTTGAACCTGGAACCTGGAACCTGGAACCTGGAACCTGGAACCTGGAAC 5100
 5101 AAGTCAGAGATTTCTCAGAACATCAAGTCTTTCATCCCACTTAA 5144
 5101 AAGTCAGAGATTTCTCAGAACATCAAGTCTTTCATCCCACTTAA 5144

RESULT 2
 LOCUS AB025611 50015 bp DNA linear PLN 27-DEC-2000
 DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone: K21B8.
 ACCESSION AB025611 BA000015
 VERSION AB025611.1 GI:4589417
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (sites)
 Kaneko, T., Katoh, T., Asamizu, E., Sato, S., Nakamura, Y., Kotani, H.
 and Tabata, S. Analysis of Arabidopsis thaliana chromosome 5. XI

Unpublished
 2 (bases 1 to 50015)
 Nakamura, Y.

Direct Submission
 Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
 Institute, Department of Plant Gene Research; 1532-3, Yana,
 Kisarazu, Chiba 252-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
 Tel: 81-438-52-3935, Fax: 81-438-52-3934)

Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
 http://www.kazusa.or.jp/kaos/cgi-bin/agd/graph.cgi?c=K21B8

Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.

The software programs used to predict genes include: Grail
 (Informatics Group, Oak Ridge National Laboratory,
 http://compbio.ornl.gov/Grail-1.3/),
 GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
 NetGene2 (S. M. Hebsgaard, et al., CBS, Technical University of
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
 SplicePredictor (Volker Brendel, Stanford University,
 http://gremlin.zool.iastate.edu/cgi-bin/sp.cgi).

Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
 http://genome.wustl.edu/eddy/tRNAscan-SE/).

This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.

The 5' clone is F6I13 and the 3' clone is MOK9.

FEATURES source	Location/Qualifiers	CDS	QGFHAHTLIFRIWVPSKRLV" complement(join(48713..49236,49497..49565)) /note="gb AAD15517.1 gene id:K21B8.8 similar to unknown protein" /pseudo /codon_start=1 /evidence=not_experimental BASE COUNT 16169 a 8857 c 7700 g 17289 t ORIGIN
	1..50015 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /strain="Columbia" /db_xref="taxon:3702" /chromosome="5" /clone="K21B8" /clone_lib="Mitsui TAC" 2444..2620 /note="gene id:K21B8.1 unknown protein" /pseudo /codon_start=1 /evidence=not_experimental join(21353..21804,21890..21991) /note="gene id:K21B8.2 pir T08458- similar to unknown protein" /pseudo /codon_start=1 /evidence=not_experimental join(23764..23928,24015..24077,24176..24247,24333..24440, 24538..24612,24681..24860,24985..25074,25178..25300, 25812..25931,26160..26276,26495..26551,26639..26713, 26798..26893) /note="gb AA62923.1 gene_id:K21B8.3" /codon_start=1 /evidence=not_experimental /product="serine/threonine protein kinase SOS2" /protein_id="BAA98146.1" /db_xref="GI:8978255" /translation="MTKMRVRVKYEVGRTIGESTFAKVKPARNTDGDNVAIKMAK STLKNRMVDQIKREISIMKIVRHPNIVRLYEVLAISPSKIYIVLFVFTGGELFDRIVH GKLESESKYFQQLVDAVAHCHGVHRDLKPNLLDNTGNLKVSDFGLSALPQ EGVELLRCTCTPNYVAPEVLSCGYDGSADLWSCGVILFVILAGVLPSETDLPL YKINAEFSCTPWFSAEVKFLIHRILDPNPKTRIQIQGKKDPWRFLNVPYIRAREE EYVNDIRAFVPGIGSVVAENRDEGFLMNNAPFEMITLSSGLNLSALFRRQDF VKQTPVRSRREPSILIANIEAVANSNGFKSRNFKRLRLEGLSSIKAGQLAVIBY EVAPSLFMDVRKAAGETIEYHKFYKKLCSKLENIWRATEGIPKSEILRTITP" complement(28006..33201) /note="contains similarity to Tall1-like non-LTR retroelement protein gene_id:K21B8.4" /pseudo /codon_start=1 /evidence=not_experimental complement(join(33610..34901,35547..37507,37633..38034)) /note="gene_id:K21B8.5" /pseudo /codon_start=1 /evidence=not_experimental /product="non-LTR retroelement reverse transcriptase-like" complement(join(38377..38751,38917..39776)) /note="gene_id:K21B8.6" /pseudo /codon_start=1 /evidence=not_experimental /product="Tall1-like non-LTR retroelement protein-like" complement(join(44026..44296,44491..44768)) /note="gb AAD56332.1 gene_id:K21B8.7 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAA98147.1" /db_xref="GI:8978256" /translation="MDPKLSERLFBELGTIEGKKVYAVFHWIKDIPVLIDDLHLK KLENSQFGLIQVMDHLNSVNMHILTRQLVTKPEFLMWTFAQPKMRLALIVLEG ILCPTIITVILPQVNMHSDLDDEFIKFPWGRSFLLTVRSVKTRKLYTVNSQGSIAI		
FEATURES source	Location/Qualifiers	CDS	QGFHAHTLIFRIWVPSKRLV" complement(join(48713..49236,49497..49565)) /note="gb AAD15517.1 gene id:K21B8.8 similar to unknown protein" /pseudo /codon_start=1 /evidence=not_experimental BASE COUNT 16169 a 8857 c 7700 g 17289 t ORIGIN
	1..50015 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /strain="Columbia" /db_xref="taxon:3702" /chromosome="5" /clone="K21B8" /clone_lib="Mitsui TAC" 2444..2620 /note="gene id:K21B8.1 unknown protein" /pseudo /codon_start=1 /evidence=not_experimental join(21353..21804,21890..21991) /note="gene id:K21B8.2 pir T08458- similar to unknown protein" /pseudo /codon_start=1 /evidence=not_experimental join(23764..23928,24015..24077,24176..24247,24333..24440, 24538..24612,24681..24860,24985..25074,25178..25300, 25812..25931,26160..26276,26495..26551,26639..26713, 26798..26893) /note="gb AA62923.1 gene_id:K21B8.3" /codon_start=1 /evidence=not_experimental /product="serine/threonine protein kinase SOS2" /protein_id="BAA98146.1" /db_xref="GI:8978255" /translation="MTKMRVRVKYEVGRTIGESTFAKVKPARNTDGDNVAIKMAK STLKNRMVDQIKREISIMKIVRHPNIVRLYEVLAISPSKIYIVLFVFTGGELFDRIVH GKLESESKYFQQLVDAVAHCHGVHRDLKPNLLDNTGNLKVSDFGLSALPQ EGVELLRCTCTPNYVAPEVLSCGYDGSADLWSCGVILFVILAGVLPSETDLPL YKINAEFSCTPWFSAEVKFLIHRILDPNPKTRIQIQGKKDPWRFLNVPYIRAREE EYVNDIRAFVPGIGSVVAENRDEGFLMNNAPFEMITLSSGLNLSALFRRQDF VKQTPVRSRREPSILIANIEAVANSNGFKSRNFKRLRLEGLSSIKAGQLAVIBY EVAPSLFMDVRKAAGETIEYHKFYKKLCSKLENIWRATEGIPKSEILRTITP" complement(28006..33201) /note="contains similarity to Tall1-like non-LTR retroelement protein gene_id:K21B8.4" /pseudo /codon_start=1 /evidence=not_experimental complement(join(33610..34901,35547..37507,37633..38034)) /note="gene_id:K21B8.5" /pseudo /codon_start=1 /evidence=not_experimental /product="non-LTR retroelement reverse transcriptase-like" complement(join(38377..38751,38917..39776)) /note="gene_id:K21B8.6" /pseudo /codon_start=1 /evidence=not_experimental /product="Tall1-like non-LTR retroelement protein-like" complement(join(44026..44296,44491..44768)) /note="gb AAD56332.1 gene_id:K21B8.7 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAA98147.1" /db_xref="GI:8978256" /translation="MDPKLSERLFBELGTIEGKKVYAVFHWIKDIPVLIDDLHLK KLENSQFGLIQVMDHLNSVNMHILTRQLVTKPEFLMWTFAQPKMRLALIVLEG ILCPTIITVILPQVNMHSDLDDEFIKFPWGRSFLLTVRSVKTRKLYTVNSQGSIAI		

Dd	22601	TGTAATCCCAACCGACTATAGGACTATCAAGAAGTTTTTCAAACTTTTTTAAAAAGATCT	22660
Qy	901	TGAACCTTTAAAGCAATCCCAACCGACTATAGAACCATAGTCTCTTCCCTGTTGATGA	960
Dd	22661	TGAACCTTTAAAGCAATCCCAACCGACTATAGAACCATAGTCTCTTCCCTGTTGATGA	22720
Qy	961	AGCTCTTCTCGTCCGCGGAAAAATCTAGGCAATAAAGCCCTCTTCAACATCACTAGTAT	1020
Dd	22721	AGCTCTTCTCGTCCGCGGAAAAATCTAGGCAATAAAGCCCTCTTCAACATCACTAGTAT	22780
Qy	1021	ATTGAACGTCACCATCTTTTGTGCAATGCTTTGTGAATGAACCGTCGATAAAACCGTGTT	1080
Dd	22781	ATTGAACGTCACCATCTTTTGTGCAATGCTTTGTGAATGAACCGTCGATAAAACCGTGTT	22840
Qy	1081	ATCACATTACGCCAAATTTTCCCTAGTGTGTTTCCAAACTCTCAGGAATCCCTTATCGA	1140
Dd	22841	ATCACATTACGCCAAATTTTCCCTAGTGTGTTTCCAAACTCTCAGGAATCCCTTATCGA	22900
Qy	1141	ACTTTTTTATATACACTTTGTAGCAATTGAAAAGTACTTTGTATGCTTTGTCTTAAACT	1200
Dd	22901	ACTTTTTTATATACACTTTGTAGCAATTGAAAAGTACTTTGTATGCTTTGTCTTAAACT	22960
Qy	1201	TAGACATCCCTGTCTCTCGTTTTTTTGAACCTTGCTTGCACTTAAATGAAGTTAAAAAT	1260
Dd	22961	TAGACATCCCTGTCTCTCGTTTTTTTGAACCTTGCTTGCACTTAAATGAAGTTAAAAAT	23020
Qy	1261	TTCTAGTTAAAAATAGAAAATTTTACTAATTTGAGTTTCAATTAATCATAGTCTAGATAAT	1320
Dd	23021	TTCTAGTTAAAAATAGAAAATTTTACTAATTTGAGTTTCAATTAATCATAGTCTAGATAAT	23080
Qy	1321	TTGAAAAAATTAATAATTAATTTTGAAAAACTATATATGTTTATTTTAAAAAATAATTTA	1380
Dd	23081	TTGAAAAAATTAATAATTAATTTTGAAAAACTATATATGTTTATTTTAAAAAATAATTTA	23140
Qy	1381	CTAAATTTGATPAAGTATATPAGATTTAGTTTTTTTTTTTTTTTTTTTTTAAATTTGAAAACTC	1440
Dd	23141	CTAAATTTGATPAAGTATATPAGATTTAGTTTTTTTTTTTTTTTTTTTTTAAATTTGAAAACTC	23200
Qy	1441	ACAATTTAAATTTGAAAGATTCAAATGCAATGTTAGTTTTTTTAAAGTTTAAATCAACCA	1500
Dd	23201	ACAATTTAAATTTGAAAGATTCAAATGCAATGTTAGTTTTTTTAAAGTTTAAATCAACCA	23260
Qy	1501	AAACGATTAATTGACCCGAACGACTAATTCAGGTCGTATACGGGTACAGTCAAAATAACCCG	1560
Dd	23261	AAACGATTAATTGACCCGAACGACTAATTCAGGTCGTATACGGGTACAGTCAAAATAACCCG	23320
Qy	1561	ACCCGAAATGCTCAACGGATCTCGACGAGCTTCGCAATGCAATTCACGGCTTTCCGGA	1620
Dd	23321	ACCCGAAATGCTCAACGGATCTCGACGAGCTTCGCAATGCAATTCACGGCTTTCCGGA	23380
Qy	1621	ATCGCATCAGAGCCTTCCCTTCTCTCTACGCCCTCTTTCATCAACCTTCTCTGCGAATC	1680
Dd	23381	ATCGCATCAGAGCCTTCCCTTCTCTCTACGCCCTCTTTCATCAACCTTCTCTGCGAATC	23440
Qy	1681	CAATTTCTGTAATTCAGATTCCTTTTCGACTACGCCCAATCGCAATTCAGCCGTAACGA	1740
Dd	23441	CAATTTCTGTAATTCAGATTCCTTTTCGACTACGCCCAATCGCAATTCAGCCGTAACGA	23500
Qy	1741	ATTTTATATATGATATGATCTTGATCTCTTACTCTTTCGCTGCTTTCCCTTGATCGAATTT	1800
Dd	23501	ATTTTATATATGATATGATCTTGATCTCTTACTCTTTCGCTGCTTTCCCTTGATCGAATTT	23560
Qy	1801	ATGTTGATTTATGCTCGGAATCAAAATTAATGATACAGAGTAATTTTGTATATATGAT	1860
Dd	23561	ATGTTGATTTATGCTCGGAATCAAAATTAATGATACAGAGTAATTTTGTATATATGAT	23620
Qy	1861	GTGTAGAGCTAAATTTAGCAAAACAATPACGTAAAGTAATTTGAAATCGAAATCAATAACGT	1920
Dd	23621	GTGTAGAGCTAAATTTAGCAAAACAATPACGTAAAGTAATTTGAAATCGAAATCAATAACGT	23680
Qy	1921	TTAAGGAAAGAGGTTTTTACTAAGTCTCTGAAATAATCTGATAGCTTTGCGTCAATG	1980
Dd	23681	TTAAGGAAAGAGGTTTTTACTAAGTCTCTGAAATAATCTGATAGCTTTGCGTCAATG	23740
Qy	1981	GATCAGATAAAAGTTTTGTAAAGATGACAAAGAAATGAGAGATGGGCAAGTACGAGG	2040
Dd	23741	GATCAGATAAAAGTTTTGTAAAGATGACAAAGAAATGAGAGATGGGCAAGTACGAGG	23800
Qy	2041	TTGGTCGCAACAATPAGGTGAAGGAACCTTTGCTTAAGCTTAAGTTTTCGAGGAACACAGACA	2100
Dd	23801	TTGGTCGCAACAATPAGGTGAAGGAACCTTTGCTTAAGCTTAAGTTTTCGAGGAACACAGACA	23860
Qy	2101	CTGGTGATAATGTAGCCATCAAAATTTATGGCTTAAGAGTACAATFACCTTAAGAAACAGAAATGG	2160
Dd	23861	CTGGTGATAATGTAGCCATCAAAATTTATGGCTTAAGAGTACAATFACCTTAAGAAACAGAAATGG	23920
Qy	2161	TTGATCAGGTATGTTCTCTCGATTGTTTTTTTACATGGAACAATAAGTTGTTGCGTCAATGGT	2220
Dd	23921	TTGATCAGGTATGTTCTCTCGATTGTTTTTTTACATGGAACAATAAGTTGTTGCGTCAATGGT	23980
Qy	2221	ATGATCTTTTGATTTCGTTTTTAAAGCTCTTTTACAGATAAAAGAGAGATATCTATAATGAA	2280
Dd	23981	ATGATCTTTTGATTTCGTTTTTAAAGCTCTTTTACAGATAAAAGAGAGATATCTATAATGAA	24040
Qy	2281	GATTTGTCGTACCCGAAACATAGTAGGTTGTATGAGGTATGTTGTTTTGTTTCCATGCA	2340
Dd	24041	GATTTGTCGTACCCGAAACATAGTAGGTTGTATGAGGTATGTTGTTTTGTTTCCATGCA	24100
Qy	2341	TCTGCGAAATTTTATCTCTGAAAGTGTTTTTTGGATCATTTGTTCTCTGTTGTTTTTTTGTG	2400
Dd	24101	TCTGCGAAATTTTATCTCTGAAAGTGTTTTTTGGATCATTTGTTCTCTGTTGTTTTTTTGTG	24160
Qy	2401	ATTTTCCCGATGATGAGTTGTCGAGTCTCTCGAAATATATATAGTTTGGAGTTTGTG	2460
Dd	24161	ATTTTCCCGATGATGAGTTGTCGAGTCTCTCGAAATATATATAGTTTGGAGTTTGTG	24220
Qy	2461	ACAGGAGGAGACTCTTTGATAGAAATTGACCGAACTTCCATCTTGTAGGACGCTCCA	2520
Dd	24221	ACAGGAGGAGACTCTTTGATAGAAATTGACCGAACTTCCATCTTGTAGGACGCTCCA	24280
Qy	2521	TTAGTTAAACCTCTCTACTTAAATTTTAAATATATGAAATCTTTTCATCGAGTTCAATA	2580
Dd	24281	TTAGTTAAACCTCTCTACTTAAATTTTAAATATATGAAATCTTTTCATCGAGTTCAATA	24340
Qy	2581	AGGAGGCTTTGAAGAAAGTGAGTCTCGGAAATACTTTCAACAGCTTGTAGATCTGTTGC	2640
Dd	24341	AGGAGGCTTTGAAGAAAGTGAGTCTCGGAAATACTTTCAACAGCTTGTAGATCTGTTGC	24400
Qy	2641	TCATTGTCATGCAAGGGTGTTTTACACCGTGAACCTTAAAGGTAAGACCTGTTTTGTTT	2700
Dd	24401	TCATTGTCATGCAAGGGTGTTTTACACCGTGAACCTTAAAGGTAAGACCTGTTTTGTTT	24460
Qy	2701	ACCAATATTTCTCAGAAATCTCACCTGCTTCCAAATCCAGACTTGGATATTTTGTGTCGC	2760
Dd	24461	ACCAATATTTCTCAGAAATCTCACCTGCTTCCAAATCCAGACTTGGATATTTTGTGTCGC	24520
Qy	2761	TATGTTATGTTATCTAGCCAGAAAAATCTTTTACTCGATACAAAATGGAATCTGAAGTTT	2820
Dd	24521	TATGTTATGTTATCTAGCCAGAAAAATCTTTTACTCGATACAAAATGGAATCTGAAGTTT	24580
Qy	2821	CGGAATTCGGACTCAGTGCATTCGCCTCAGGAAGTAAGTCTCTTATCTCTGCTCAGCAG	2880
Dd	24581	CGGAATTCGGACTCAGTGCATTCGCCTCAGGAAGTAAGTCTCTTATCTCTGCTCAGCAG	24640
Qy	2881	TCTGCTTACGTGCTCAATTAACCTTGTATATATCTCAATCAGGAGTAGAATCTTCGGTAC	2940
Dd	24641	TCTGCTTACGTGCTCAATTAACCTTGTATATATCTCAATCAGGAGTAGAATCTTCGGTAC	24700
Qy	2941	CACATGTGAACTCCGAACTATGTAGCTCCAGAGTACTTAGTGACAGGGTTACGATGG	3000
Dd	24701	CACATGTGAACTCCGAACTATGTAGCTCCAGAGTACTTAGTGACAGGGTTACGATGG	24760
Qy	3001	TTACAGCAGCTGATATTTTGGTCTTGGGGGTTATCTTTTCGTTATATTTGGCTGGATTTT	3060
Dd	24761	TTACAGCAGCTGATATTTTGGTCTTGGGGGTTATCTTTTCGTTATATTTGGCTGGATTTT	24820

QY 3061 ACCATTTTCCGAGCGGATCTTCCAGGGTGTGACAGAAAGTAAGTACATATCTTTGG 3120
Db 24821 ACCATTTTCCGAGCGGATCTTCCAGGGTGTGACAGAAAGTAAGTACATATCTTTGG 24880
QY 3121 GAAGAAATCATGAATTCCTTCTCATGGCTTTTGCAAAACCGCTTTATTTGATTTGGTTTCG 3180
Db 24881 GAAGAAATCATGAATTCCTTCTCATGGCTTTTGCAAAACCGCTTTATTTGATTTGGTTTCG 24940
QY 3181 AATTTTCAACCTTAGTTTGTAGCTTTTACATTTTACATTTGTTATTTACAGATAAATCGACAGAGT 3240
Db 24941 AATTTTCAACCTTAGTTTGTAGCTTTTACATTTTACATTTGTTATTTACAGATAAATCGACAGAGT 25000
QY 3241 TTCTTTGTCCACCGTGGTTTTCCGAGAGTGAAGTTTTTAAATACATAGGATATCTTGACC 3300
Db 25001 TTCTTTGTCCACCGTGGTTTTCCGAGAGTGAAGTTTTTAAATACATAGGATATCTTGACC 25060
QY 3301 CCAATCCCAAAACAGTGAGTATTTTGGCTTTCTCTCTAGCTATCAGGTTTTTGGTGAT 3360
Db 25061 CCAATCCCAAAACAGTGAGTATTTTGGCTTTCTCTCTAGCTATCAGGTTTTTGGTGAT 25120
QY 3361 ATTTAAGTCTCTAGTAAATATATCTGTTTATCTATTTATTTGTTCTCAATTAGACAGCGT 3420
Db 25121 ATTTAAGTCTCTAGTAAATATATCTGTTTATCTATTTATTTGTTCTCAATTAGACAGCGT 25180
QY 3421 ATTCAAATTCAGGAATCAAGAAATCAAGAAAGATCCITGGTTTCAGATTAAATTAATGTGCCATTACGA 3480
Db 25181 ATTCAAATTCAGGAATCAAGAAAGATCCITGGTTTCAGATTAAATTAATGTGCCATTACGA 25240
QY 3481 GCAAGGAAGAAGAAAGTGAATTTGGATGATATTCGTGCAGTTTTTGTGGAATTTGAG 3540
Db 25241 GCAAGGAAGAAGAAAGTGAATTTGGATGATATTCGTGCAGTTTTTGTGGAATTTGAG 25300
QY 3541 GTTTGTCTTCGCTTCATTTATATGCTCTTGTGCTGCTGCTCAATTCATTTAAATGTTAAGAT 3600
Db 25301 GTTTGTCTTCGCTTCATTTATATGCTCTTGTGCTGCTCAATTCATTTAAATGTTAAGAT 25360
QY 3601 CTCTTAGGAACGTTTGGATGACCAAGAAAGTGTGTTGCTACAGGATAGAACAAAATAGT 3660
Db 25361 CTCTTAGGAACGTTTGGATGACCAAGAAAGTGTGTTGCTACAGGATAGAACAAAATAGT 25420
QY 3661 AGCATGTGTTAGTTACCAACCTGTAAACCTGTTCTTTTCAATTCGCGCAACCATAG 3720
Db 25421 AGCATGTGTTAGTTACCAACCTGTAAACCTGTTCTTTTCAATTCGCGCAACCATAG 25480
QY 3721 ACCTTAGGAACGTTAGATTACAGAGAAATTCCTATTTCTCGACCAAAAACCTGACGA 3780
Db 25481 ACCTTAGGAACGTTAGATTACAGAGAAATTCCTATTTCTCGACCAAAAACCTGACGA 25540
QY 3781 AAATCCAGAATACCCCTAGGGCTAAATPACAAATGTTCAATGTACCTATCAATATATATCTCC 3840
Db 25541 AAATCCAGAATACCCCTAGGGCTAAATPACAAATGTTCAATGTACCTATCAATATATATCTCC 25600
QY 3841 TGCTAAGATTGCTTAGCTTTGGTAAGCCTAGATATATATAGATACCGTAATTTCTTAAT 3900
Db 25601 TGCTAAGATTGCTTAGCTTTGGTAAGCCTAGATATATATAGATACCGTAATTTCTTAAT 25660
QY 3901 GCATATTTAGAAATGTTTTTACTTAAACCTCAGGCTCTTGCTCTTCTAAACCTGTACTTCAA 3960
Db 25661 GCATATTTAGAAATGTTTTTACTTAAACCTCAGGCTCTTGCTCTTCTAAACCTGTACTTCAA 25720
QY 3961 TTGTTAAACTTAAACCTCAGTATCTGTCTTAGCTTAAAGTACTTTTTTACTTTGTTTTTCAAT 4020
Db 25721 TTGTTAAACTTAAACCTCAGTATCTGTCTTAGCTTAAAGTACTTTTTTACTTTGTTTTTCAAT 25780
QY 4021 AAGTTGACCTGTCAATTTGCACTTTGTTCAAGGGCAGTTATGTAGCGGAGAAATGTAGAGAG 4080
Db 25781 AAGTTGACCTGTCAATTTGCACTTTGTTCAAGGGCAGTTATGTAGCGGAGAAATGTAGAGAG 25840
QY 4081 AAATGATGAAGGGCCCTGATGATGATGCTTTTGGATGCTTTTGGATGATTTACCTTATCAAGGCTT 4140
Db 25841 AAATGATGAAGGGCCCTGATGATGATGCTTTTGGATGCTTTTGGATGATTTACCTTATCAAGGCTT 25900
QY 4141 AAATTTATCTGCACATTTTGACAGGGGACAGGTAGTACCTGATTTTCTTATTTACTGCTCAT 4200

Db 25901 AAATTTATCTGCACATATTTTGACAGCGACAGGTAGTACCTGATTTTCTAATCTGTCAT 25960
QY 4201 AGAGATCTCCATTTCCGAATAAAGAAATGTCGGTAGCATCTATTCTTCAGACTGCCCGTTT 4260
Db 25961 AGAGATCTCCATTTCCGAATAAAGAAATGTCGGTAGCATCTATTCTTCAGACTGCCCGTTT 26020
QY 4261 TGAATGCTCTTATGATGCTGTGTTCTTAGTTTGTATTAATAAATCAATTAAGTTTCATTTAGATGA 4320
Db 26021 TGAATGCTCTTATGATGCTGTGTTCTTAGTTTGTATTAATAAATCAATTAAGTTTCATTTAGATGA 26080
QY 4321 TTGGTTCCATGGCATTAGTAGATACAAATGGAATCCAAAATGTTCTTCGCATTTAGTGGC 4380
Db 26081 TTGGTTCCATGGCATTAGTAGATACAAATGGAATCCAAAATGTTCTTCGCATTTAGTGGC 26140
QY 4381 TGATCTCTTTGATCTCGCAGGATTTTGTAAAGGCAACCCGTTTTTCTCGAAGGGA 4440
Db 26141 TGATCTCTTTGATCTCGCAGGATTTTGTAAAGGCAACCCGTTTTTCTCGAAGGGA 26200
QY 4441 ACCTAGTGAGATAATTTGCTTAACATTTAGGCTGTAGCGAACTCAATGGGTTTTTAAGTCTCA 4500
Db 26201 ACCTAGTGAGATAATTTGCTTAACATTTAGGCTGTAGCGAACTCAATGGGTTTTTAAGTCTCA 26260
QY 4501 TACACGAAACCTTCAAGGTAAAGAAATTCCTAGCATATTTACACTTATCACAGAGATTATGCA 4560
Db 26261 TACACGAAACCTTCAAGGTAAAGAAATTCCTAGCATATTTACACTTATCACAGAGATTATGCA 26320
QY 4561 TTATTTTAAACCTCTCAACTGTTTAAAGCGATGTGTAGATGATTAAGATTGCAAAATTTT 4620
Db 26321 TTATTTTAAACCTCTCAACTGTTTAAAGCGATGTGTAGATGATTAAGATTGCAAAATTTT 26380
QY 4621 AAATTTAGTTTATATCTCTGGCGTTTCAAAAACGAAAGTCTCTAGTGTGAATTTATCAATTTT 4680
Db 26381 AAATTTAGTTTATATCTCTGGCGTTTCAAAAACGAAAGTCTCTAGTGTGAATTTATCAATTTT 26440
QY 4681 AATGTTTAGCAGAGAGTACAAATCTGTATGATTTGCTTAAAGTCTGTCTCAACAGACAAG 4740
Db 26441 AATGTTTAGCAGAGAGTACAAATCTGTATGATTTGCTTAAAGTCTGTCTCAACAGACAAG 26500
QY 4741 CTCGAGGGAATTAATCTCGATCAAGGCGGACAGTTAGCTGTTGTGTATAGAGTAAATTTAT 4800
Db 26501 CTCGAGGGAATTAATCTCGATCAAGGCGGACAGTTAGCTGTTGTGTATAGAGTAAATTTAT 26560
QY 4801 GCTTGTGTGATTTGAATTAATTAAGTTTCTCTTGTCTTCAAGTTTAAAGGTTATCTAGCAATTTG 4860
Db 26561 GCTTGTGTGATTTGAATTAATTAAGTTTCTCTTGTCTTCAAGTTTAAAGGTTATCTAGCAATTTG 26620
QY 4861 AAATTAACCTACATGCAGATTTTACAGGTGGGCAACCTCGTTTTCATGTTAGACGTAAGA 4920
Db 26621 AAATTAACCTACATGCAGATTTTACAGGTGGGCAACCTCGTTTTCATGTTAGACGTAAGA 26680
QY 4921 AAGGCTGCTGGTGAATCTCTTGAATATCAAGGTTTATAAATATATATATATATATATATATAT 4980
Db 26681 AAGGCTGCTGGTGAATCTCTTGAATATCAAGGTTTATAAATATATATATATATATATATATAT 26740
QY 4981 AGTTGCATCATTTACTGTTGTGCGGATTAAGTGAATTTTGTGTTTGTGTGTATGCGAGTTTC 5040
Db 26741 AGTTGCATCATTTACTGTTGTGCGGATTAAGTGAATTTTGTGTTTGTGTGTATGCGAGTTTC 26800
QY 5041 TACAAGAACTATGTTTCCGAACCTGGAACATATAATATGAGGGCAACAGAGAAATACCA 5100
Db 26801 TACAAGAACTATGTTTCCGAACCTGGAACATATAATATGAGGGCAACAGAGAAATACCA 26860
QY 5101 AAGTCAGAGATTCTCAGAAACAATCACTGTTTGTATCCCACTTAA 5144
Db 26861 AAGTCAGAGATTCTCAGAAACAATCACTGTTTGTATCCCACTTAA 26904

RESULT 3
LOCUS
DEFINITION
ACCESSION

F6113
Arabidopsis thaliana
16228 bp
BAC F6113.
linear
PLN 26-JUL-2000
AF262044

AF262044.1	GI:7682804																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
------------	------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

QY	961	AGCTCTTCTCGTGGCGGCGAAAATCTAGGCCATAAAGAGCTCTTCAACATCACCTAGTAT	1020
DB	12662	AGCTCTTCTCGTGGCGGCGAAAATCTAGGCCATAAAGAGCTCTTCAACATCACCTAGTAT	12721
QY	1021	ATTGACCGTGACCAATCTTTTGGACCATTTGTCCTTTGTGAATGAACCGTTCGATAAACCGTGT	1080
DB	12722	ATTGACCGTGACCAATCTTTTGGACCATTTGTCCTTTGTGAATGAACCGTTCGATAAACCGTGT	12781
QY	1081	ATCACTTACGCGCAAAATTTTTCCCTAGTGTGTTGTTTCCAACTCTCACGAATCCTTATCGA	1140
DB	12782	ATCACTTACGCGCAAAATTTTTCCCTAGTGTGTTTCCAACTCTCACGAATCCTTATCGA	12841
QY	1141	ACTTTTATATATATCACTTTGTAGCATTTGAAAGTATCTTTGTATGCTTTGTCTTAAACT	1200
DB	12842	ACTTTTATATATATCACTTTGTAGCATTTGAAAGTATCTTTGTATGCTTTGTCTTAAACT	12901
QY	1201	TAGACATCCTTGTCTCTGTTTGTGTTTGTGAACTTGTGACTTAAATGAAGTTAAAAAT	1260
DB	12902	TAGACATCCTTGTCTCTGTTTGTGTTTGTGAACTTGTGACTTAAATGAAGTTAAAAAT	12961
QY	1261	TTGTAGTTAAAAATAGAAAAATTTTACTAAATTCAGTTTCGATTAAATCATATGCTAGATAAT	1320
DB	12962	TTGTAGTTAAAAATAGAAAAATTTTACTAAATTCAGTTTCGATTAAATCATATGCTAGATAAT	13021
QY	1321	TTGAAAAAAATTTAAATAAATTTTGAAAAATCATATATGTTTATTTTTTAAAAATAAATTA	1380
DB	13022	TTGAAAAAAATTTAAATAAATTTTGAAAAATCATATATGTTTATTTTTTAAAAATAAATTA	13081
QY	1381	CTAAATTTGTAAGTGATTTAGATTAGTTTATTTTCTTTTCTTTTAAATTTGAAACCTC	1440
DB	13082	CTAAATTTGTAAGTGATTTAGATTAGTTTATTTTCTTTTCTTTTAAATTTGAAACCTC	13141
QY	1441	ACAATTTAAATTTGAAAGATTCAAATGCAATGTTAGTTTAAAAAGTTTAAATCAACCCA	1500
DB	13142	ACAATTTAAATTTGAAAGATTCAAATGCAATGTTAGTTTAAAAAGTTTAAATCAACCCA	13201
QY	1501	AACGGATAATTGACCGGAACGACTAATTCAGTTCGTATACGGGTACAGTCAAAATAACCG	1560
DB	13202	AACGGATAATTGACCGGAACGACTAATTCAGTTCGTATACGGGTACAGTCAAAATAACCG	13261
QY	1561	ACCGAAATGCTCAACGGATCTCGACGGAGGTTCCGATCGACGATTCAGGCTTCGGGA	1620
DB	13262	ACCGAAATGCTCAACGGATCTCGACGGAGGTTCCGATCGACGATTCAGGCTTCGGGA	13321
QY	1621	ATCGCATCAGAGCCTTCCTTCTCTCGTACGCTCTTTTCATCAACCCCTTCTCTCGGAATC	1680
DB	13322	ATCGCATCAGAGCCTTCCTTCTCTCGTACGCTCTTTTCATCAACCCCTTCTCTCGGAATC	13381
QY	1681	CAATTCGGTATTTACGATCTCTTTTCGACTACGCCCAATCGCAATTCAGCCGTACGA	1740
DB	13382	CAATTCGGTATTTACGATCTCTTTTCGACTACGCCCAATCGCAATTCAGCCGTACGA	13441
QY	1741	ATTTTATATATTCATTATGATCTTGATCTTACCTTTTCGTTCTCGTTCTTGATCGAATTT	1800
DB	13442	ATTTTATATATTCATTATGATCTTGATCTTACCTTTTCGTTCTCGTTCTTGATCGAATTT	13501
QY	1801	ATGTTGATTTATGCTCGGAAATCAAATAATATGATACAGAGTAAATTTTGTTATATATGAT	1860
DB	13502	ATGTTGATTTATGCTCGGAAATCAAATAATATGATACAGAGTAAATTTTGTTATATATGAT	13561
QY	1861	GTGTAGAGCTAAAATTAGCAAAACAATACGTAAAGTAAATGAAATCGAAATCATAAACGT	1920
DB	13562	GTGTAGAGCTAAAATTAGCAAAACAATACGTAAAGTAAATGAAATCGAAATCATAAACGT	13621
QY	1921	TTAAGGAAGAGGTTTTTACTAGTCTCTGAAATATCTGATTTGATAGCTTTGTGTCATATG	1980
DB	13622	TTAAGGAAGAGGTTTTTACTAGTCTCTGAAATATCTGATTTGATAGCTTTGTGTCATATG	13681
QY	1981	GATCAGATAAAGTTTGTGTAAAGAAATGACAAAGAAATGAGAGAGTGGCGAAGTACGAGG	2040
DB	13682	GATCAGATAAAGTTTGTGTAAAGAAATGACAAAGAAATGAGAGAGTGGCGAAGTACGAGG	13741
QY	2041	TTGGTCGCAATAGGTGAAGGAACCTTTTGCTAAGGTTAAGTTTTCGGAGGAACACAGACA	2100

Db	13742	TTGGTCGACAAATAGGTGAAGAACCTTTGCTAAGTTAAGTTTGGAGGAACACAGACA	13801
Qy	2101	CTGGTGAATAAGTAGCCATCAAAATTTATGGCTAAGAGTACAATACTTAAGAACAGAAATGG	2160
Db	13802	CTGGTGAATAAGTAGCCATCAAAATTTATGGCTAAGAGTACAATACTTAAGAACAGAAATGG	13861
Qy	2161	TTGATCAGGTATGTTCTGGATGTTTTTTTACATGGAACTAAGGTTGTTGGTCAATGGT	2220
Db	13862	TTGATCAGGTATGTTCTGGATGTTTTTTTACATGGAACTAAGGTTGTTGGTCAATGGT	13921
Qy	2221	ATGATCTTTTGATTTCTGTTTAAAGCTCTTTTACAGATAAAAGAGAGATATCTATAATGAA	2280
Db	13922	ATGATCTTTTGATTTCTGTTTAAAGCTCTTTTACAGATAAAAGAGAGATATCTATAATGAA	13981
Qy	2281	GATTGTTCTGCACCCGAAACATAGTAGGTTGTATAGGATATGTTTGTGTTTCCATGCA	2340
Db	13982	GATTGTTCTGCACCCGAAACATAGTAGGTTGTATAGGATATGTTTGTGTTTCCATGCA	14041
Qy	2341	TCTCGGAATTTTATCTCTGAAGTGTTTTTGCATCATTTGTTCTTCTGTGTTTTTTTTGTG	2400
Db	14042	TCTCGGAATTTTATCTCTGAAGTGTTTTTGCATCATTTGTTCTTCTGTGTTTTTTTTGTG	14101
Qy	2401	ATTTTCCCGATAGGTTGTGGCGAGTCCTTCGAAAAATATATATAGTTTGGAGTTGTG	2460
Db	14102	ATTTTCCCGATAGGTTGTGGCGAGTCCTTCGAAAAATATATATAGTTTGGAGTTGTG	14161
Qy	2461	ACAGGAGGAGCTCTTTGATAGAAATGTACGGAACCTCCATPACTTGTAGGACGCGTCCA	2520
Db	14162	ACAGGAGGAGCTCTTTGATAGAAATGTACGGAACCTCCATPACTTGTAGGACGCGTCCA	14221
Qy	2521	TTAGTTAAAACCTCTACTTAATTTTTTAAATATAGAACTTTTCATGCGAGTTCATAA	2580
Db	14222	TTAGTTAAAACCTCTCTACTTAATTTTTTAAATATAGAACTTTTCATGCGAGTTCATAA	14281
Qy	2581	AGGAGGCTTTGAAGAAAGTAGTGCTCGAAAAATACTTTCAACAGCTGTAGATGCTGTGC	2640
Db	14282	AGGAGGCTTTGAAGAAAGTAGTGCTCGAAAAATACTTTCAACAGCTGTAGATGCTGTGC	14341
Qy	2641	TCATGTGACTGCAAGGGTGTTTACACCGTGACTAAAGGTAAAGACGTGTTTTTGTTT	2700
Db	14342	TCATGTGACTGCAAGGGTGTTTACACCGTGACTAAAGGTAAAGACGTGTTTTTGTTT	14401
Qy	2701	ACCAATATTCCTCAGAAATATCTCAGTCGGTTCCAAATCAGACTTGATATTTTGTGCGC	2760
Db	14402	ACCAATATTCCTCAGAAATATCTCAGTCGGTTCCAAATCAGACTTGATATTTTGTGCGC	14461
Qy	2761	TATGTTATGTTATCTAGCCAGAAAAATCTTTTACTCGATACAAATGGAATCTGAAGGTTT	2820
Db	14462	TATGTTATGTTATCTAGCCAGAAAAATCTTTTACTCGATACAAATGGAATCTGAAGGTTT	14521
Qy	2821	CGAATTCGACATCAGTGCAATTCGCTCAGGAAGTAAAGTGCTTTATCTCTGCTCAGCAG	2880
Db	14522	CGAATTCGACATCAGTGCAATTCGCTCAGGAAGTAAAGTGCTTTATCTCTGCTCAGCAG	14581
Qy	2881	TCCTGTTACTGCTGTCATTAACTTGTTATATATCTCAATCAGGAGTAGAACTTCTCCGTC	2940
Db	14582	TCCTGTTACTGCTGTCATTAACTTGTTATATATCTCAATCAGGAGTAGAACTTCTCCGTC	14641
Qy	2941	CACATGTGGAACCTCCGAACTATGTAGCTCCAGAGGTACTTATGTGACAGAGGTTACGATGG	3000
Db	14642	CACATGTGGAACCTCCGAACTATGTAGCTCCAGAGGTACTTATGTGACAGGTTACGATGG	14701
Qy	3001	TTCAGCAGCTGATATTTTGGCTTTGCGGGTTATTTCTTTTCGTTATATATGGCTCGATATT	3060
Db	14702	TTCAGCAGCTGATATTTTGGCTTTGCGGGTTATTTCTTTTCGTTATATATGGCTCGATATT	14761
Qy	3061	ACCTTTTTCCGAGACGATCTTCCAGGGTTGTACGAAAAAGTAAAGTAACTATCTTTCCG	3120
Db	14762	ACCTTTTTCCGAGACGATCTTCCAGGGTTGTACGAAAAAGTAAAGTAACTATCTTTCCG	14821
Qy	3121	GAAGAAATCATGAAATCCCTTTGTCATGGCTTTTGTCAAAACCGTTTATTTGATTTGGTTTGC	3180

Db 14822 GAAGAAATCATGAATTCCTTGTGTCATGGCTTTTGTCAAAACCGTTTATTGATTGGTTTGC 14881

Qy 3181 AATTTCACCCCTTAGCTTTTGTAGCTTTTACACATGTTTATTACAGATAAATGCGACAGAT 3240

Db 14882 AATTTCACCCCTTAGCTTTTGTAGCTTTTACACATGTTTATTACAGATAAATGCGACAGAT 14941

Qy 3241 TTTCTTGTGCCACCGTGGTTTCCGACAGAAGTGAAGTATTTTAAATACATAGATACCTTGACC 3300

Db 14942 TTTCTTGTGCCACCGTGGTTTCCGACAGAAGTGAAGTATTTTAAATACATAGATACCTTGACC 15001

Qy 3301 CCAATCCAAACAGTGAATTTTGTCTTGTCTCTCCTAGCTATCAGGTTTTTGTGAT 3360

Db 15002 CCAATCCAAACAGTGAATTTTGTCTTGTCTCTCCTAGCTATCAGGTTTTTGTGAT 15061

Qy 3361 ATTAAATGTTCTAGTAATTATATCTGTTTATCTATTAATGTTTCTCAATTAGACACGGT 3420

Db 15062 ATTAAATGTTCTAGTAATTATATCTGTTTATCTATTAATGTTTCTCAATTAGACACGGT 15121

Qy 3421 ATTCAAATTCAGGAATCAAGAAAGATCCTTGGTTTCAGATTAAATTAATGTGCCCTATACGA 3480

Db 15122 ATTCAAATTCAGGAATCAAGAAAGATCCTTGGTTTCAGATTAAATTAATGTGCCCTATACGA 15181

Qy 3481 GCAAGGAAGAAGAGTGAATTTGGATGATATTCGTGCAGTATTTTGTGGAATTGAG 3540

Db 15182 GCAAGGAAGAAGAGTGAATTTGGATGATATTCGTGCAGTATTTTGTGGAATTGAG 15241

Qy 3541 GTTTGTGCTTGCCTTCATATATGCTTTTGTGCTGCTCAATTCATTAATGTTAGAT 3600

Db 15242 GTTTGTGCTTGCCTTCATATATGCTTTTGTGCTGCTCAATTCATTAATGTTAGAT 15301

Qy 3601 CTCTTAGGAACGTTTGGATGACCAAGAAGTCTTTGTCTACAGATAGAACAAAATAGT 3660

Db 15302 CTCTTAGGAACGTTTGGATGACCAAGAAGTCTTTGTCTACAGATAGAACAAAATAGT 15361

Qy 3661 AGCATGTTTGTAGTACAAACCTGTAACTGCTTCTTTAFTCAATTCGCAACCATAG 3720

Db 15362 AGCATGTTTGTAGTACAAACCTGTAACTGCTTCTTTAFTCAATTCGCAACCATAG 15421

Qy 3721 ACCTTAGGAACGTTTGTAGTACAAACCTGTAACTGCTTCTTTAFTCAATTCGCAACCATAG 3780

Db 15422 ACCTTAGGAACGTTTGTAGTACAAACCTGTAACTGCTTCTTTAFTCAATTCGCAACCATAG 15481

Qy 3781 AAATCCAGATACCCCTAGGCTAAATACATGTTTCATGTTACCTATCAATATATATCTCC 3840

Db 15482 AAATCCAGATACCCCTAGGCTAAATACATGTTTCATGTTACCTATCAATATATATCTCC 15541

Qy 3841 TGCTAAGATGCTTGTAGTACGCTTAGATATATATAGATACCGTAATTTCTAAT 3900

Db 15542 TGCTAAGATGCTTGTAGTACGCTTAGATATATATAGATACCGTAATTTCTAAT 15601

Qy 3901 GCATATTTAGAATGTTTACTTAACTCAGGCTCTTCTCTTAAACCTGTACTTCAA 3960

Db 15602 GCATATTTAGAATGTTTACTTAACTCAGGCTCTTCTCTTAAACCTGTACTTCAA 15661

Qy 3961 TTGTTAACTAAACCTCAGTATCTGTCTTAGCTTAAAGTACTTTTACTTGTTTTCAAT 4020

Db 15662 TTGTTAACTAAACCTCAGTATCTGTCTTAGCTTAAAGTACTTTTACTTGTTTTCAAT 15721

Qy 4021 AAGTTGACCTGTCAATTTGCACTTGTTCACAGGCGAGTTATGTAGCGAGATGTAGAGAG 4080

Db 15722 AAGTTGACCTGTCAATTTGCACTTGTTCACAGGCGAGTTATGTAGCGAGATGTAGAGAG 15781

Qy 4081 AAATGATGAAGGGCCCTGATGATGATGCTTTGAGATGATTAACCTATCACAGGCTT 4140

Db 15782 AAATGATGAAGGGCCCTGATGATGATGCTTTGAGATGATTAACCTATCACAGGCTT 15841

Qy 4141 AAATTTATCTGCATATTTGACAGGCGACAGTACTACCTGATTTTCTTACTTGGTCAAT 4200

Db 15842 AAATTTATCTGCATATTTGACAGGCGACAGTACTACCTGATTTTCTTACTTGGTCAAT 15901

Qy 4201 AGAGATCTCATTTCGAATAAAGAAATGTCGGTAGCATCTAATCTTCAGACTGCCCGTTT 4260

Db 15902 AGAGATCTCATTTCGAATAAAGAAATGTCGGTAGCATCTAATCTTCAGACTGCCCGTTT 15961

Qy 4261 TGACTGCCTTATGATGCTGTGTTCTTAGTGTGTTTATAAATAACTATATAAGTTCAATTAGATGA 4320

Db 15962 TGACTGCCTTATGATGCTGTGTTCTTAGTGTGTTTATAAATAACTATATAAGTTCAATTAGATGA 16021

Qy 4321 TTGTTTGCATGGCATTAGTAGATACAAATGGAATCCAAATGTTTCCCTGCATATTGATGGC 4380

Db 16022 TTGTTTGCATGGCATTAGTAGATACAAATGGAATCCAAATGTTTCCCTGCATATTGATGGC 16081

Qy 4381 TGAATCCTTTCATCTCGCAGGATTTTGTAAAGGCCAAACCCGTTTGTCTTCGAAGGGA 4440

Db 16082 TGAATCCTTTCATCTCGCAGGATTTTGTAAAGGCCAAACCCGTTTGTCTTCGAAGGGA 16141

Qy 4441 ACCTAGTGAGATTAATGCTTAACATTGAGGCTGTAGCGAACTCAATGGTGTTTAAGTCTCA 4500

Db 16142 ACCTAGTGAGATTAATGCTTAACATTGAGGCTGTAGCGAACTCAATGGTGTTTAAGTCTCA 16201

Qy 4501 TACACGAAACTTCAAGGTAAACGAATTC 4527

Db 16202 TACACGAACTTCAAGGTAAACGAATTC 16228

RESULT 4
AC006258/c
LOCUS
DEFINITION
Arabidopsis thaliana BAC F18G18 from chromosome V near 60.5 cM,
complete sequence.
ACCESSION
AC006258
VERSION
AC006258.1 GI:4063735
KEYWORDS
HTG.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 110469)
Huang, E.N., Dedhia, N., de la Bastide, M., Habermann, K., Matero, A.,
Preston, R., Shekher, M., Rodriguez, M., Nascimento, L., Spiegel, L.A.,
Schutz, K., Shah, R., Swaby, I., Vill, M.D., O'Shaughnessy, A.,
Parnell, L.D. and McCombie, W.R.
Arabidopsis thaliana BAC F18G18 from chromosome V near 60.5 cM
Unpublished
2 (bases 1 to 110469)
Huang, E.N., Dedhia, N., de la Bastide, M., Habermann, K., Matero, A.,
Preston, R., Shekher, M., Rodriguez, M., Nascimento, L., Spiegel, L.A.,
Schutz, K., Shah, R., Swaby, I., Vill, M.D., O'Shaughnessy, A.,
Parnell, L.D. and McCombie, W.R.
Direct Submission
Submitted (28-DEC-1998) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, PO Box 100, Cold Spring
Harbor, NY 11724, USA
BAC F18G18 is assigned to YAC C1C2E3 and maps to near 60.5 cM on
the Lister & Dean RI map. Position 1 of F18G18 is oriented toward
the telomere and position 110469 is oriented toward the centromere.
For more information on the mapping, sequencing and annotation of
F18G18, please see
http://www.cshl.org/arabweb/F18G18-tilepage.html. A graphic view
of our annotation is also available at this url. Gene models are
built with exons predicted by GenScan
(http://CCR-081.mit.edu/GENSCAN.html), MZEF
(http://www.cshl.org/genefinder) and GRAIL
(http://compbio.ornl.gov/tools/index.shtml) and with splice sites
predicted by NetPlantGene
(http://www.cbs.dtu.dk/netgene/cbsnetpgene.html). Alternate exons
not used in building the gene models are presented on the web pages
associated with F18G18. Genes are numbered according to the scheme
BAC gene number. Typically, these numbers progress from 1 upwards
as one moves from position 1 of the BAC. Protein sequences encoded
by the genes are assigned to a functional category with the aid of
similarity searches and comparison to the Prosite
(http://expasy.hcuge.ch/sprot/prosite.html) and Pfam
(http://pfam.wustl.edu/) libraries. A description of these
categories can be found at

Db 105643 TAAAGTACATAATTAACCTACCAATATTGAATTGATTTTAAATA 105605

RESULT 6

AC007399/c

LOCUS AC007399 105223 bp DNA linear PLN 26-APR-1999

DEFINITION Arabidopsis thaliana BAC F14123 from chromosome V near 69 cM, complete sequence.

ACCESSION AC007399

VERSION AC007399.1 GI:4680765

KEYWORDS HTG.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1 (bases 1 to 105223)

AUTHORS Habermann, K., Nascimben, L., Vil, M.D., Matero, A., Rodriguez, M., Shah, R., Swaby, I., Shekher, M., O'Shaughnessy, A., Huang, E.N., Spiegel, L.A., Schutzi, K., Parnell, L.D., Preston, R.R., See, L.H., Dedhia, N.N. and McCombie, W.R.

TITLE Arabidopsis thaliana BAC F14123 from chromosome V near 69 cM

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 105223)

AUTHORS Habermann, K., Nascimben, L., Vil, M.D., Matero, A., Rodriguez, M., Shah, R., Swaby, I., Shekher, M., O'Shaughnessy, A., Huang, E.N., Spiegel, L.A., Schutzi, K., Parnell, L.D., Preston, R.R., See, L.H., Dedhia, N.N. and McCombie, W.R.

TITLE Direct Submission

JOURNAL Submitted (26-APR-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724

COMMENT BAC F14123 maps to near 69 cM on the Lister & Dean RI map and is assigned to YAC C1C12F8. Position 1 of F14123 is oriented toward the north telomere centromere and position 105223 is oriented toward the centromere. For more information on the mapping sequencing and annotation of F14123, please see <http://www.cshl.org/arabweb/F14123-titlepage.html>. A graphic view of our annotation will also be available at this url. Gene models are built with exons predicted by GenScan (<http://CCR-081.mit.edu/GENSCAN.html>), MZEF (<http://www.cshl.org/genefinder>) and GRAIL (<http://combio.ornl.gov/tools/index.shtml>) and with splice sites predicted by NetPlantGene (<http://www.cbs.dtu.dk/netpene/cbsnetpgene.html>). Genes are numbered according to the scheme BAC.gene.number. Typically, these numbers progress from 1 upwards as one moves from position 1 of the BAC. Hypothetical proteins are those having no EST matches and similarity only to other hypothetical proteins; predicted proteins have EST matches but cannot be assigned a function based on current composition and nomenclature within the nr protein database. Protein sequences encoded by the genes are assigned to a functional category with the aid of similarity searches and comparison to the Prosite (<http://expasy.hcuge.ch/sprot/prosite.html>) and Pfam (<http://pfam.wustl.edu/>) libraries. A description of these categories can be found at <http://muntjak.mips.biochem.mpg.de/arabi/>. Genomic repeats are typically located by TBLASTX analysis and an attempt is made to classify the function of each repeat as either transposon, putative microsatellite, LINE, direct repeat, centromeric repeat, etc.

If you have any questions or confirmatory or contradictory evidence concerning the annotation of F14123, please direct email to Larry Parnell at parnell@cshl.org.

FEATURES

source Location/Qualifiers

1..105223

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosomes="V"

/map="near 69 cM"

```

BASE COUNT      34317 a 19530 c 18267 g 33109 t
ORIGIN
Query Match      8.9%; Score 457.6; DB 8; Length 105223;
Best Local Similarity 92.5%; Pred. No. 2.1e-62;
Matches 492; Conservative 0; Mismatches 39; Indels 1; Gaps 1;

QY 1 TTTTGGCGGGAATACTCGGGTTTACGTTTGGCGGGAATACTCGGTTCGTTTACGTTT 60
Db TTTTGGCGGGAATACTCGGGTTTACGTTTGGCGGGAATACTCGGTTCGTTTACGTTT 60900

QY 61 TGGCGGGAATACTCGGGTTTACGTTTGGCGGGAATACTCGGGTTTACGTTT 120
Db TGGCGGGAATACTCGGGTTTACGTTTGGCGGGAATACTCGGGTTTACGTTT 60841

QY 121 CGGGAATACTCAGGAATTATGTTTTGGTGGAATACTCAGGTTTACGTTTCTCAAT 180
Db CGGGAATACTCAGGTTTATGTTTTGGTGGAATACTCAGGTTTACGTTTCTCAAT 60781

QY 181 TTCATTGCTGTATATTAAGAAATTTGGAAAAATATTAATTTTAAATTTGGTTTGA 240
Db TTCATCGCTGTATATTAAGAAATTTGGAAAAATATTAATTTTAAATTTGGTTTGA 60721

QY 241 TGTGTTGGTTAAACCTAAATTTGGCAATTTGGTTTACGTTTGGTTTATTTCAATTT 300
Db TGTGTTGGTTAAACCTAAATTTGGCAATTTGGTTTACGTTTGGTTTATTTGAATTT 60661

QY 301 TACAAAATTTGATGGTTAATTTGATTAACCATCGGAACCAATTAACCATTAACCTAAC 360
Db TACAAAATTTGATGGTTAATTTGATTAACCATTTGAAACCAATTAACCATTAACCTAAC 60601

QY 361 TCATTTTACTCATCAACCACTGACATCACTCACTCACTCACTCACTCACTCACTCACT 420
Db TCATTTTACTCATCAACCACTGACATCACTCACTCACTCACTCACTCACTCACTCACT 60541

QY 421 TCAAAAATTTCAACTCAATAGGTTTCATGGTTTGGTTTGGTTTGGTTTGGTTTGGTT 480
Db TCAAAAATTTCAACTCAATAGGTTTCATGGTTTGGTTTGGTTTGGTTTGGTTTGGTT 60481

QY 481 TTGACCCATTTTGACACCCCTACATATGATCATTAAGTTAATAATCAAAAATTT 532
Db TTGACCCATTTTGACACCCCTACAGGAAGTATTAGATAATGATTAAATAATTT 60429

RESULT 7
AC007627/c
LOCUS
DEFINITION      Genomic Sequence For Arabidopsis thaliana Clone F15F15, Chromosome
ACCESSION      AC007627
VERSION        AC007627.3 GI:6449503
KEYWORDS       HTG.
SOURCE         Arabidopsis thaliana (thale cress)
ORGANISM       Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 157000)
Spiegel, L.A., Nascimento, L.U., de la Bastide, M., Vil, D.M.,
Huang, E.N., Preston, R.R., Matero, A., Shah, R., O'Shaughnessy, A.,
Rodriguez, M., Shekher, M., Swaby, I., Schutz, K., See, L.H.,
Habermann, K., Dedhia, N.N. and McCombie, W.R.
Genomic Sequence For Arabidopsis thaliana Clone F15F15, Chromosome
V, Complete Sequence
Unpublished
2 (bases 1 to 157000)
McCombie, W.R.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS
Submitted (21-MAY-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
JOURNAL
REFERENCE
AUTHORS
3 (bases 1 to 157000)
McCombie, W.R.

Direct Submission
Submitted (18-NOV-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
COMMENT
This clone overlaps F14I23 (Accession number AC007399) from bases
1-48320 (matches from 56939-105223 on F14I23). This clone also
overlaps from base 146390-15700 with T24G3 (Accession number
AC006192). The overlapping region is from base 35-10685 on T24G3.
FEATURES
Location/Qualifiers
source
1..157000
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="v"
/map="near 69cm"
/clone_lib="IGF"
3150..3162
misc_feature
/note="There is a difference in the number of A's called.
The positive clones call 12 and the negative clones call
11."
124150..124230
misc_feature
/note="The assembly is single stranded with single
chemistry."
125300..125340
misc_feature
/note="The assembly is single stranded with single
chemistry"
BASE COUNT      51913 a 27270 c 26614 g 51203 t
ORIGIN
Query Match      8.9%; Score 457.6; DB 8; Length 157000;
Best Local Similarity 92.5%; Pred. No. 2e-62;
Matches 492; Conservative 0; Mismatches 39; Indels 1; Gaps 1;

QY 1 TTTTGGCGGGAATACTCGGGTTTACGTTTGGCGGGAATACTCGGTTCGTTTACGTTT 60
Db TTTTGGCGGGAATACTCGGGTTTACGTTTGGCGGGAATACTCGGTTCGTTTACGTTT 3961

QY 61 TGGCGGGAATACTCGGGTTTACGTTTGGCGGGAATACTCGGGTTTACGTTT 120
Db TGGCGGGAATACTCGGGTTTACGTTTGGCGGGAATACTCGGGTTTACGTTT 3902

QY 121 CGGGAATACTCAGGAATTATGTTTTGGTGGAATACTCAGGTTTACGTTTCTCAAT 180
Db CGGGAATACTCAGGTTTATGTTTTGGTGGAATACTCAGGTTTACGTTTCTCAAT 3842

QY 181 TTCATTGCTGTATATTAAGAAATTTGGAAAAATATTAATTTTAAATTTGGTTTGA 240
Db TTCATCGCTGTATATTAAGAAATTTGGAAAAATATTAATTTTAAATTTGGTTTGA 3782

QY 241 TGTGTTGGTTAAACCTAAATTTGGCAATTTGGTTTACGTTTGGTTTATTTCAATTT 300
Db TGTGTTGGTTAAACCTAAATTTGGCAATTTGGTTTACGTTTGGTTTATTTGAATTT 3722

QY 301 TACAAAATTTGATGGTTAATTTGATTAACCACTGGAACCAATTAACCATTAACCTAAC 360
Db TACAAAATTTGATGGTTAATTTGATTAACCACTTGAACCAATTAACCATTAACCTAAC 3662

QY 361 TCATTTTACTCATCAACCACTGACATCACTCACTCACTCACTCACTCACTCACTCACT 420
Db TCATTTTACTCATCAACCACTGACATCACTCACTCACTCACTCACTCACTCACTCACT 3602

QY 421 TCAAAAATTTCAACTCAATAGGTTTCATGGTTTGGTTTGGTTTGGTTTGGTTTGGTT 480
Db TCAAAAATTTCAACTCAATAGGTTTCATGGTTTGGTTTGGTTTGGTTTGGTTTGGTT 3542

QY 481 TTGACCCATTTTGACACCCCTACATATGATCATTAAGTTAATAATCAAAAATTT 532
Db TTGACCCATTTTGACACCCCTACAGGAAGTATTAGATAATGATTAAATAATTT 3490

RESULT 8

```


AC007123/c	AC007123	133195 bp	DNA	linear	PLN 20-MAR-1999
LOCUS	Arabidopsis thaliana BAC F21A20	from chromosome V near 61 cM, complete sequence.			
DEFINITION	AC007123				
ACCESSION	AC007123.1	GI:4454587			
VERSION	HTG.				
KEYWORDS	Arabidopsis thaliana (thale cress)				
SOURCE	Arabidopsis thaliana				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (bases 1 to 133195)				
AUTHORS	Spiegel,L.A., Huang,E.N., Rodriguez,M., O'Shaughnessy,A., Shekher,M., Habermann,K., Schut,K., Wil,M.D., Shah,R., Swaby,I., Matero,A., de la Bastide,M., Preston,R.R., See,L.H., Nascimento,L., Dedhia,N.N., Parnell,L.D. and McCombie,W.R.				
TITLE	Arabidopsis thaliana BAC F21A20	from chromosome V near 61 cM			
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 133195)				
AUTHORS	Spiegel,L.A., Huang,E.N., Rodriguez,M., O'Shaughnessy,A., Shekher,M., Habermann,K., Schut,K., Wil,M.D., Shah,R., Swaby,I., Matero,A., de la Bastide,M., Preston,R.R., See,L.H., Nascimento,L., Dedhia,N.N., Parnell,L.D. and McCombie,W.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-MAR-1999)	Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724			
COMMENT	BAC F21A20	is a clone from chromosome V, near 61 cM. Position 1 of F21A20 is the T7 end and is oriented toward the north telomere. Position 133195 is the Sf6 end and is oriented toward the centromere. The north neighboring clone is T21B4 and the south neighboring clone is F15A18.			
FEATURES	Location/Qualifiers				
source	1. .133195				
	/organism="Arabidopsis thaliana"				
	/mol_type="genomic DNA"				
	/cultivar="Columbia"				
	/db_xref="taxon:3702"				
	/chromosome="V"				
	/map="near 61 cM"				
misc_feature	1. .23913				
	/notes="overlap with T21B4, GenBank accession number AF007271, from position 1 to 23913"				
repeat_region	complement(120541..120797)				
	/notes="function-retrotransposon fragment; similar to TSC1 retrotransposon, GenBank accession number U65470"				
	/rpt_type="dispersed"				
BASE COUNT	43384 a 22442 c 22975 g 44394 t				
ORIGIN					
	Query Match	8.9%; Score 456.2; DB 8; Length 133195;			
	Best Local Similarity	93.5%; Pred. No. 3.4e-62;			
	Matches 487; Conservative	0; Mismatches 33; Indels 1; Gaps 1;			
Qy	1	TTTTTGGCGGAAAAATCTCGGTTACGTTTTTGGCGGAAAAATCTCGGTTACGTTTT	60		
Db	41145	TTTTTGGCGGAAAAATACGGGTTACGTTTTTGGCGGAAAAATCTCGGTTACGTTTT	41086		
Qy	61	TGGCGGAAAAATCTCGGTTACGTTTTTGGCGGAAAAATCTCGGTTACGTTTTTTG	120		
Db	41085	TGGCGGAAAAATCTCGGTTACGTTTTTGGCGGAAAAATCTCGGTTACGTTTTTG	41027		
Qy	121	CGGAAAAATCAGGATTATGTTTTTGGTGGAAAAATACGAGTTTCTCTCAAT	180		
Db	41026	CGGAAAAATCAGGTTATGTTTTTGGTGGAAAAATACGAGTTCTCTCAAT	40967		
Qy	181	TTCAATGCTTGTATATTAGAAATTTGAAAAATATTATTTTATTAATTCGTTTGA	240		
Db	40966	TTCAATGCTTGTATATTAGAAATTTGAAAAATATTATTTTATTAATTCGTTTGA	40907		
Qy	241	TGTGTTGGTTAAACCTAAATGCGATTGTTAGAGATTTTGTGTTTATTCAAATTT	300		

IOVQSFPTALKSRKEVHGYSYATTQATCVPSARRRCTPQFFFAANTCFAKIAPOS
 GEVFLHAIILFVIEYLDVERM"
 1824. .2316
 /gene="T7H20_10"
 /number=1
 2317. .2402
 /gene="T7H20_10"
 /number=1
 2403. .2479
 /gene="T7H20_10"
 /number=2
 2480. .2559
 /gene="T7H20_10"
 /number=2
 2560. .2721
 /gene="T7H20_10"
 /number=3
 2722. .2834
 /gene="T7H20_10"
 /number=3
 2835. .2918
 /gene="T7H20_10"
 /number=4
 2919. .3086
 /gene="T7H20_10"
 /number=4
 3087. .3251
 /gene="T7H20_10"
 /number=5
 3252. .3301
 /gene="T7H20_10"
 /number=5
 3302. .3406
 /gene="T7H20_10"
 /number=6
 3407. .3512
 /gene="T7H20_10"
 /number=6
 3513. .3668
 /gene="T7H20_10"
 /number=7
 4027. .5664
 /gene="T7H20_20"
 complement (join(4027. .4332,4414. .4581,4673. .4876,
 4955. .5080,5179. .5283,5518. .5664))
 /gene="T7H20_20"
 complement (join(4027. .4332,4414. .4581,4673. .4876,
 4955. .5080,5179. .5283,5518. .5664))
 /gene="T7H20_20"
 /note="similarity to putative protein At2g30530 -
 Arabidopsis thaliana, EMBL:U93215"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB82970.1"
 /db_xref="GI:7340671"
 /translation="MAYRRRGIGKPKATFKEEVDRLPPESITAVKDRSPAPSPAPD
 OPSKNETTEPKGLWGIAGKAKSVIEDKSDRSSTASQSRFSYLSDEGFKMDNPK
 LRGLDKLSSLNQIGDIFKAFEDGRLTVENKTADIIQTRKLTFRRTTGGEDNQ
 NQSYGVSSSKPQEPQMLNHEHETQLKASRDVAMATAAKALLRELKTVKADLA
 FAKERCALEEENKGRHREKSNPADEDLRLQESLLAEKARIAHENSIVAREN
 RFREIVEXHQLTWQDVYIDEGSEEVTVQSPFVSTLMTSPADRSQSPPSQEMIXEM
 LVVAESSVCASVVSVS"
 complement (4027. .4332)
 /gene="T7H20_20"
 /number=1
 complement (4333. .4413)
 /gene="T7H20_20"
 /number=1
 complement (4414. .4581)
 /gene="T7H20_20"
 /number=2
 complement (4582. .4672)
 /gene="T7H20_20"
 /number=2

/number=2
 complement (4673. .4876)
 /gene="T7H20_20"
 /number=3
 complement (4877. .4954)
 /gene="T7H20_20"
 /number=3
 complement (4955. .5080)
 /gene="T7H20_20"
 /number=4
 complement (5081. .5178)
 /gene="T7H20_20"
 /number=4
 complement (5179. .5283)
 /gene="T7H20_20"
 /number=5
 complement (5284. .5517)
 /gene="T7H20_20"
 /number=5
 complement (5518. .5664)
 /gene="T7H20_20"
 /number=6
 6555. .8036
 /gene="T7H20_30"
 6555. .8036
 /gene="T7H20_30"
 /note="similarity to COP1-interacting protein CIP8 -
 Arabidopsis thaliana, EMBL:AF162150"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB82971.1"
 /db_xref="GI:7340672"
 /translation="MESQPSNGHPMDISSNIPSSACNLGCRVVOASDLDLEIFSLC
 VDCKFLLEDFTGPAEDPLARRQTIERRRRRTTRHDSESVDGDLSSQDFTHLI
 SIARQSLSTVQASDELDDLTSSHTTPSGTRMRLFSSESDDFNFGTEGNAGFS
 LYRFHADNDTTFSAYGESDASTDRHADI FVQPDNRSDIDFDTIDPMRAGLNOWN
 SDEEDREBEGAGPSGAGTGYNYLASPSSESYSDRSDSPLELERSFQRIIERHS
 LSRNIFTGLDLDLFSYANVGDYLDERGFDELLEQLAESDNRRAAPPASVSCVNNL
 PRVITAEHVMKGLVCAICKELFSLRNFTTQLFCLHLYHAHCIVPWLARNSCPLCRY
 ELPTDDKDYEEGRGNLDVDSSESSDDGTGTEGEEYVERGESGVSRVSRGRWLFL
 AAAPVSVLGVVLAWLMLSSPHRRGIAISHSQRENTRRWLPFF"
 6555. .8036
 /gene="T7H20_30"
 /number=1
 8386. .10613
 /gene="T7H20_40"
 complement (join(8386. .8696,8788. .8840,8920. .9020,
 9134. .9208,9311. .9541,9843. .9964,10042. .10166,
 10245. .10324,10416. .10613))
 /gene="T7H20_40"
 complement (join(8386. .8696,8788. .8840,8920. .9020,
 9134. .9208,9311. .9541,9843. .9964,10042. .10166,
 10245. .10324,10416. .10613))
 /gene="T7H20_40"
 /note="similarity to hypothetical proteins - Arabidopsis
 thaliana"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB82972.1"
 /db_xref="GI:7340673"
 /translation="MIARILAAADSMEMPVAAGGSVLGTTIKIATWPKYPTMCTFL
 GILMASKYNIILPPSCRKLLGLVFSLLLPCLIFSLGQAVTLOKMLQWFFIPVNVVL
 GRISSIGFIIVASIVRPPYFKFTIIQIGVGNIGNVFLVLLAALCRUTSNPPFGDSE
 KCSIDGTAYISFGWGAILIYVYVFPAPPFGDFEENLAIKLTVDAPEQVP
 LITONPKDPSPTQDLLPVQSTEPGRGVSGKIAQIFVLYEKLKIQVQPAIVA
 SILAMILGAIPTFKLIIFNGAPLPFTDSMLGLDAMPICILLALGNGLINGPGSSK
 LQFKTTAAIILGLVLPVPGIVTVADKGLFADDDKMFREVLLOHTMTSIVLSG
 AVANLRCGCGRESAAVFWWHIFAIFSMAGNMVLYINILF"
 complement (8386. .8696)
 /gene="T7H20_40"
 /number=1
 complement (8697. .8787)
 /gene="T7H20_40"
 /number=1


```
/protein_id="AAF79388.1"
/db_xref="GI:8778380"
/translacion="MTWRLTNLPQTLGHLFTQDARNNSVFKKYTRNLFPPTSVSTPTIT
WTKTAPFVKCNDFAGHINHSLSGTGWIIRDHOGSAHAGSSIVDVHNTVPSEAEATKA
LLVAMQOQWTGKXMIQFEGDHSYLASSEFTNSIFOADTRNOPPLANLLCKNLHS"
join(6536..6622,6671..6964,6989..7075)
/notes="similar to phosphoenolpyruvate carboxykinase
(ATP)-like protein emb|CAB38935.1"
/codon_start=1
/evidence=not experimental
/product="F1504.6"
/protein_id="AAF79344.1"
/db_xref="GI:8778336"
/translacion="WISQNFYIARVFLICALEMFCYTEYLSVILYEQAIYEVKGSLLIT
SNGALTWLSGAKTDREISVLEMLLELLRMSGDYRIFILYGLLIFLDFDFDFFGLGSLN
TEIDHPTVMVNRRAVDYLNLSKILSHSGNMGKGDGVFAFFELLGATF"
complement(join(9342..9431,9588..9662,9771..9872,
10106..10186,10497..10811))
/notes="putative 50S ribosomal protein L21, chloroplast
precursor sp|P51412|RK21 ARATH; similar to ESTs
gb|T45781.1 and gb|AW004148.1"
/codon_start=1
/evidence=not experimental
/product="F1504.7"
/protein_id="AAF79387.1"
/db_xref="GI:8778379"
/translacion="MASSATLSLCSFSAHCNVNSPRSSLTLSLKPSLNLAKEPLT
GFLSPASTSRVAFATVAPFASVVEAEPTDIEAVVSDVSEVTEKAKREEIFA
VIVGSGQNDKVVVFYTKPKKKYRRNIGRQPNTRIRITGITGYEYPASPVAAGV
NL"
complement(join(11530..11670,11756..11980,12064..12231,
12303..12418,12517..12669,12759..12902,13095..13225,
13337..14132))
/notes="similar to calcium-dependent protein kinase SKS
sp|P28583|CDPK_SOYBN; similar to ESTs gb|AW728580.1,
gb|A1994248.1, and gb|A1997698.1"
/codon_start=1
/evidence=not experimental
/product="F1504.8"
/protein_id="AAF79386.1"
/db_xref="GI:8778378"
/translacion="METKPNRPPSNTVLPYQTLRLRDHVLKKGLOGQFGTYLCT
EKSTANYACKSIPKRLVCREDEYDVMBIQIMHLSHPNVVRLKGTVEDSVFVHI
VMEVCEGELFDRIIVSKGFHSEAEVKLITLIGVVEACHSLGWMERDLKPNFLFDS
PKDAKGLATDGLSVFYKPEGLIFLWLDLSLIQLVFWLFFVSMRKLKQSGFLIET
GLLCFTWIANRNVKPIREFEDDLFGQYLVDVGSPPYVAPEVLLKKGPEIDVMSAGV
ILXILSGVFPFVAETESGIFRQILQKLDLDFKDPWPTTISEAAKOLITIKMLRSPPKR
ISAHEALCPWIDEQAADPKPLDPALVSELKQFSQWNKI KQVALRVIAERLSEEEIG
GLKELFKMIDTNSGNTTPEELKAGLRVGSSELMESEIKSLMDAADIDNSGTTDYGEF
LAATLHMNMRENEIIVAAFSYFDKQSGVITIDELQSACTEFLGLCDTPLDDMIKSID
LNDGKIDRSEFTAMRKGDGVRSTMMKNLNFNTADAFGVGKSSD"
join(17932..18090,19698..19920,20052..20233)
/notes="similar to En/Spm-like transposon protein
gb|AAC14510.1"
/codon_start=1
/evidence=not experimental
/product="F1504.9"
/protein_id="AAF79345.1"
/db_xref="GI:8778337"
/translacion="MFKLSAQQLFWYAMLKAACNKYLKPSLTCGLFILSGTKVTER
VLHDFITKVKIIGFMELISIDMFKPMTLTKVHPKLRSDRRVWHPKFGFIGAMDCTH
VPAKVRQDOQRVNRNKNKCSMNILGKYFVDSGYLRCGYLGPYQSRVHPHFQNO
APPYNYKEFNRLHFSRVSRIEFGV"
20230..20508
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/product="F1504.10"
/protein_id="AAF79346.1"
/db_xref="GI:8778338"
/translacion="MKRKMEDNASRARYDVRITRKLIVATMALHNFVRKSNIPDPDFE
ANCEQEGNHQPSLNEEVEQEDQDMIDSRQYMEGIRDDIAMNLWNHR"
```

CDS

```
complement(join(21563..21662,21848..22122,22212..24551,
24643..24717,24801..24989,25089..25340,25419..25941,
26020..26227,26306..26741))
/notes="unknown protein; similar to EST gb|N37451.1"
/codon_start=1
/evidence=not experimental
/product="F1504.11"
/protein_id="AAF79385.1"
/db_xref="GI:8778377"
/translacion="MEQSAAPQPTMTPLPYLPSSVEASRDDLOCIGTWMIVBPKPVGPF
LCGSIPLVADNSFPASPTALLPSQETLVLTSPSSSSSSSSSSSSSVLALCLKLFFIC
CSVUTAPRYQMLPMETDLNLPILITFPDNLVLAAYKSRITGDISKEANVTNSLKS
KCALAVLRNNGPCTKNVDVILKIPKIPYSKARVSIAPVQVTLVLPNGEDV
EEGKILRRNNQPKTNVDESILFLNFAMHSVRMEACDIPPHRETEKRSSSALP
AGNSHDNA PDDLRLDPAGSKQKODGFCCEKKKKKNKAGVPPVRKQI0ISEKIKS
SGSEKHSRGSGNEFLRVLPQPHNFMILGSLLLFNSNKYVAVSLHLDWSEKVPV
LTWLEALVNLVMAVPELATCYHENGIVQVYELLTDDIIFLKGISEDGTAPAPHPVV
QONGLAIVRLQSNCKEDPGAYWLYKSAGDEQLQFLDLSIIKNSHSSVNDSSSPS
LHSGSDSFSIGNLIYRVHRLSVNVNDRNKCARFLTQCLNCLDADPHLWVRAY
AHEQFARLLINSDESDLTPEFSGVQREVKITDLEBALDPVTIADHENTVTFSDEK
FTEDHSVINPLVSVRPKLEANVSLKELLHSDSPSDHTEGSVANSDDTSLDLCG
LCQTTSPISSKLSAINHYSOAIKSLRWLQLOSQVDAFHDLDPFSKSCGDPDC
IEVCDIRKWLPTSKLDRKLNVLVLLGESVLSLGEAYKEDQIHLQALNTVELACSLYG
SMPQKEETLTVSSMKNKSLQSKFHERTQVEDEAKSGPSDISEELSTRFLWAKY
WMLVGLIYVQPHILKQGLSRRTKTNHLMQSEVVEVKEVORLKKLITEVSCASC
LVNCSKDRASGSSSSSSSSSARTVPHSRKHNKLOSNVASKVSDVEDERVNF
KVNKRKEEDTSGTGAVRLEQNESNKETPGAKGGIPKYLKSGKSTDDAESNLL
AALNCEYETRALQELFNSCSEFQSVLRKKGVNCELGRNLKSLKLNKAEDAFADAI
VAPKEVCDHNTVILINCNLGHRRALAEVUPKI EALELHRA FENAYOKALGTAKLEY
SKSLRYMAAKTSLSVATAEASSVSNLKVUVYTOANTYLRFGMLLANEDTTAARE
QKILNTHDSSDCKSSDLRREVLASDAIREALALYESIGELRKEAEAFAYLOIA
RYHKOCCLGFLETERQSPKPSNVIOAKOYALLADRNWQKSMDFYGPENLPFSLF
TILIERASLSTSVNFWQLNFMLESALSILLEGRIHISKTVAESLRTEDPKLYTKFMAQ
LQMLKRLMALSLPSEKANSQTCGRSGSDSKLRELYKTSKSTNLCDLNAMEALLIQV
ISCSLVNKCDSVVKHRIVFGGGSVNVLSG"
join(27313..27574,27603..27721,27735..28416,28568..28584,
29183..29237,29327..29469,29559..29835,29998..30455,
30581..30652,30731..30820)
/notes="hypothetical protein"
```

CDS

```
Query Match      8.8%; Score 453.4; DB 8; Length 158096;
Best Local Similarity 81.6%; Pred. No. 8.9e-62;
Matches 551; Conservative 0; Mismatches 116; Indels 8; Gaps 2;

Qy 1 TTTTTCGCGGAAAAATCTCGGTTTACGTTTTTGGCGGAAAAATCTCGTTTACGTTTT 60
Db TTTTTCGCGGAAAAATCTCGGTTTACGTTTTTGGCGGAAAAATCTCGGTTTACGTTTT 52628

Qy 61 TGGCGGAAAAATCTCGGTTTACGTTTTTTCGAGAAATCAGGGTTTACGTTTTTGG 120
Db TGGCGGAAAAATCTCGGTTTACGTTTTTGGCGGAAAAATCTCGGTTTACGTTTTTGG 52627

Qy 121 CGGAAAAATCAGGATTATGTTTTTGGTGGAAAAATACGAGTTTACTTTTTCTCAAT 180
Db CGGAAAAATCAGGATTATGTTTTTGGTGGAAAAATACGAGTTTACTTTTTCTCAAT 52509

Qy 181 TTCAATGCTGTATATTAAGAAATTTGAAAAATAATTAATTTTATAAATGGTTTAGA 240
Db TTATCGCTGTATATTAAGAAATTTGAAAAATAATTAATTTTATAAATGGTTTAGA 52449

Qy 241 TGTGTTGGTTAAACCTAAATTCGATTGTTTAGAGATTTTAGTGGTTTTTATCAATTT 300
Db TGTGTTGGTTAAACCTAAATTCGATTGTTTAGAGATTTTAGTGGTTTTTATCAATTT 52448

Qy 301 TACAAAAATTTGATGGGTTAATTTGGATAAACCATGAAACCATTAACCATCAACCTAAC 360
Db TACAAAAATTTGATGGGTTAATTTGGTAAACCATTAACCATTAACCATCAACCTAAC 52388

Qy 361 TCATTTTACTCATCAACCAATTCATCAACTCATTTGACTCATCAACTCATTTGAG 420
Db TCATTTTACTCATCAACCAATTCATCAACTCATTTGACTCATTAACCTCATTTGAG 52328

Qy 421 TCAAAAAATTTCAACTCATTTAGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGAGTT 480
```

CDS

CDS

```

Db      52268 TC AAAATTTTCAACTCATTAGGTTTCATGGTTGAGTTGAGTTGAGTTGATTCATCATGATTT 52209
QY      481 TTGACCCATTTTGACACCCCTACATATGATC-----ATAAGTAAATAATCAAAAATTA 533
Db      52208 TTGACCCATTTTGACACCCCTAGTTATATACCAATCCAAAAGAAACGTTTGCAATGAA 52149
QY      534 CTATTGTATACTTTTAAACGGAATGTTTGTAGTTTTCATTACGTTTATTATATATAAAA 593
Db      52148 TCATCCCAAAACTTAATGAGAAATCTGACTAAAACAGGTAGCTTTTAAATTTGAGACGA 52089
QY      594 AAACATAATGCAAAAGTACTAATGTATAGTATTATTTTATTTTATTTTAAATAATGCAAAATAATT 653
Db      52088 CGAGAAAAAGAAAAAGAAAAAGAAATAGGTATCTTTTCTAGAGTTTGAACAAACCA 52029
QY      654 ACTGTAATACATTTT 668
Db      52028 AACTGTGTTTTTTTTT 52014

RESULT 11
LOCUS   AC011809
DEFINITION Arabidopsis thaliana chromosome I BAC F6A14 linear PLN 30-OCT-2002
complete sequence.
ACCESSION AC011809
VERSION   AC011809.2 GI:6579253
KEYWORDS HTG.
SOURCE   Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE
AUTHORS Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
Altai, H., Araujo, R., Huizar, I., Rowley, D., Buehler, E., Dunn, P.,
Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.S.,
Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
Unpublished
JOURNAL
REFERENCE
AUTHORS Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
Altai, H., Nguyen, M., Lam, B., Southwick, A., Bei, O., Buehler, E.,
Chin, C., Chlou, J., Choi, E., Dunn, P., Gonzalez, A., Howng, B., Kim, C.,
Koo, T., Lee, J.M., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S.,
Mukharshi, N., Pham, P., Sakano, H., Schwartz, J., Shinn, P.,
Thavari, A., Toriumi, M., Vaysberg, M., Walker, M., Yu, G., Ecker, J.,
Theologis, A. and Davis, R.W.
Direct Submission
Submitted (15-OCT-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
(bases 1 to 108767)
REFERENCE
AUTHORS Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
Altai, H., Araujo, R., Huizar, I., Rowley, D., Buehler, E., Dunn, P.,
Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S.,
Luros, S., Schwartz, J., Shinn, P., Toriumi, M., Vysotskaia, V.,
Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.
Direct Submission
Submitted (15-DEC-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
(bases 1 to 108767)
REFERENCE
AUTHORS Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,
Altai, H., Nguyen, M., Lam, B., Southwick, A., Ecker, J., Theologis, A.
and Davis, R.W.
Direct Submission
Submitted (22-JAN-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Dec 15, 1999 this sequence version replaced gi:6041764.
e-mail for correspondence: arab@sequence.stanford.edu

```

Genes with similarity to proteins in the databases are described as 'putative', '-like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/section/index.html>), GENSCAN (Chris Burge, <http://genomic.stanford.edu/~chris/GENSCAN.html>), Fexa (V.Solovyev & A.Salamov, Sanger Centre, <http://genomic.sanger.ac.uk/>), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/NetPlantGene.html>).

FEATURES

source

```

1. 108767
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosomes="I"
/clone="F6A14"
4134..6402
/gene="F6A14.1"
join(4134..4785,4919..5062,5155..5307,5397..5512,
5621..5788,5881..6111,6229..6402)
/gene="F6A14.1"
/note="calcium-dependent protein kinase 1, gb|S46283, the
cloned cDNA for this protein starts at nucleotide 4160 of
this BAC (gb|AC011809), 24 bp before the predicted start
codon."
/codon_start=1
/product="calcium-dependent protein kinase 1"
/protein_id="AAF27092.1"
/db_xref="GI:6730697"
/translation="MGNCNACVRPDSKESKPSKPKPNDRKLNFPAGDTRSPAPI
RVLDVPMNQTSITLPEHPNVVKLKASVEDNENHVMELCEGLDFRIVARGHYT
VIEDVREVAIMSTLPEHPNVVKLKASVEDNENHVMELCEGLDFRIVARGHYT
EAAATAVATIAVVMCHSGVHMDLAKENFLPANKENSLKAIKIDGLSVFRPGD
KEDATVIGSPVYMAPELVKEDYGVGVDSAGVILYLLCGVPPFWAEVTEQGVALL
RGVLDFKDPFWQISEASAKSLVKQMLDPTKRLTAQVIAHPWIOAKKAPFUGD
I VRSRLKQSPMNRFKKVLRLVIAEHLSTQREVIKMFSLMDDDKDKITYPELFG
LQKVGSLGEPFIRKMEVADVGNGFDYGFVAVIHLOKIEDELKFLAPFDFK
DASTYIELDREALDELGEPAASVLSIDIMREVDTKDGRINIDYDEVTMMKAGTDWR
KASRYSRERFKSLSLINLMKDGSLHLHDALTQGTVPV"
complement(6629..9070)
/gene="F6A14.2"
complement(join(6629..7529,7607..8160,8258..8475,
8380..9070))
/gene="F6A14.2"
/note="Similar to peptide transport proteins"
/codon_start=1
/protein_id="AAF27093.1"
/db_xref="GI:6730698"
/translation="MEVEKTEKNIAEDDSKIYRGKVMFPFIIGNTEFKLGIVGSS
SNLIYLTTFVNMKSTIAKVVNIYGTSTFGTIVAAFLCDVSGFRKYLSTAMIACF
LGSVAMDLTAHPLHPACAKSIGVCGNPGSIGOIMLAGAMVLIVAGGIRPCNL
PGAQDFDPTKTEKGEGLIESFFNVEFTFQAQMSLTLLVYQSNVNSIGLAPAI
IMLGLCITFGAGSKLYVKVKAQSGSPIHSITRVVVAIKRKLKVPGNELNYIASDF
KMSKLHTEQFRLDKSAIQDDKLNKDGSPDAWKLCSCMQQVEVKCVIRVLPVLF
SAALFYLAIQDTTITFQSLQDRRLGEGSQIPAGSYTFVLMGMTTIFITDRLV
VFLRRYTGQGGITQLQKVGAGLFCITSMVSAIVEOYRRKVALTKETLGLAPRK
ATSSMSGWMLIPOLVMIGADALAGVOMEFYKQPPENMRSPAGSLYYCGIGLASYL
STFLASAVHDTTTEGSGSWLPEDLNKGRLEVFYFLVAGMTNLNLAFLYVSHWRYK
DVVAKDKMDKTSASFDKSV"
complement(10694..14124)
/gene="F6A14.3"
complement(join(10694..10741,10837..10929,11012..11109,
11244..11328,11532..11642,11809..11866,11962..12080,
12177..12245,12346..12420,12498..12644,12724..12840,
12920..13037,13131..13366,13960..14124))
/gene="F6A14.3"
/note="Putative isochorismate synthase"
/codon_start=1
/protein_id="AAF27094.1"
/db_xref="GI:6730699"

```

```

/translaton="MNGCEADHKAPLCTVETRLSTVSPAAATERLITAVSDLKSP
PPSSGTVRLQVPIEQKIGALDWLHAONEILPSFFSRSDSGRPDLQLQFSDNGSS
DHPVSVAGLSAVFFRDLDFFSHDDWRSIRRLFSLSKSPILRIAYGGRLFPDTGKIAYE
WEHFSYFVTVQVDFEFGSSMLAATVAMDNLSTWLEAIEALQETMLQVSSVIM
RLRRESLGVIVSNKHVPSGAYPAVNALEIIKOKHSLKVLVILARSRIITDIDI
DPIATLARLOCEGDAYQFCLQPGAPAFINTPTFRHKLKVCSEALAAATPRGD
SKVREMIERDLTSPKDDLEFSTVRENIREKLTICDRVVVPHKSVKYLARVOHLY
SOLAGOLKREDDENILTAHPTPAVCGCPVEEARLLIKQIESFDRGMVAGPIGFRGG
GESEFVGISALVEKGLIYAGTGVISGNSPSSSEWNELEKISOFTKSLSEHSAAL
QPIN"
complement (18661..20376)
/gene="F6A14.5"
complement (join(18661..19422,19510..19623,19837..20376))
/gene="F6A14.5"
/notes="Hypothetical protein"
/codon_start=1
/protein_id="AAF27095.1"
/db_xref="GI:6730700"
/translaton="MDRAKEENRLKSLKIKKDFILOTOYNOLMAKHNEPTKQS
KGHODKGEDEDEKUNERELVLSLGRRLNSEVPSGSKNEKNKDVAEADRYND
DNKSSITQGLSMGLYKALSNPEKLEIDHNQETMSLEISNNKRIQSONFGPKNDG
DHEDEBILPNLVKTRVMDNDGQWKRYGKIAGNCPPRAYRYCTTIAASCPRKQ
VQCESDMSLITSEYTHNPLPMSATMASATSAASMLSGASSSSAAADLHLG
NFSLSGNNITPKKTHFLQSPSSGHPVTLDLTSSSOQPFSLMLNRSPSPNVS
RNSYPSNTLNFNTNMLNMGWGGNPSDQVRAAYGNINTHOQSPVHKIIOTRTAGS
SPDPGFRSSSHSPQINLDHIGIKNIISHQVPSLPARTIKAITIDRSPFQALATALSS
IMGDLKIDHNVRENAKSP"
23743..24942
/gene="F6A14.6"
23743..24942
/gene="F6A14.6"
/notes="Unknown protein; Location of ESTs TASG082,
gb|Z18037 and 15861277, gb|AA720219"
/codon_start=1
/protein_id="AAF27096.1"
/db_xref="GI:6730701"
/translaton="MTLLEAISNAVANDKVDQSDVPIPLSHDGI PANLKPLENPN
LGTLPNIPISGWSIQSDEVIDLKGKFSKLBKLDKDTNGFVKDEVMKLFQLEKIG
EKVGLISPAESEMVLSQAGSEIQILVKEKGNFLMGDRVSGVLVKGCLISLEWELVEILI
SNLSVDRHSYSYLVNLEKQSDILCVIKASDLDGATELLSILKFLCPSKEAIST
MAKVREWESQAMLAIEKVENTELSKVAEEASILLMVHADGFSTSELCLHYLLAS
RNYDEVFASAGAKNGNMGSPFIRYLSKWKMKVEMFPQAGPCPKAASLGLKLCNNV
PELTDITIKCLGLLIDENFSTLVLYSLDHLBELSIAKVADGLASESKLSCFVANVVSLL
KLGAARN"
26772..28802
/gene="F6A14.7"
join(26772..27077,27487..27567,27648..28677,28771..28802)
/notes="Unknown protein; Location of EST G12B177,
gb|W43460"
/codon_start=1
/protein_id="AAF27097.1"
/db_xref="GI:6730702"
/translaton="MYHVQRIYNGKEVVVVISKIEESDVVDSLSSIGNAAVYTSGIVE
TQNLKHEVDSDIYQSEVQTDQSVASVPDSSLSESEKIQOEIAAIVTQAAVRYGL
ENKANSVDDTHYGLNKGTLGNAFAOKLLASSPNVLPLSLDNDSSNISWLENWASCP
WKPVPQPKASLRSKOKKASNPQIVAEAPKSVKVPVSNLNSNSVAQSSBLE
KPKRSKYSTSSVSPLESDMPQVDLEKVRGLRGLRVNPNVENSIIQPLVPOIAVE
KPNGLSEESNADEDEKEDVAVTVQPEELIOTHTPLGTNLESLDSTLVNQIEESE
NVNAEKEDVKERTKQNHKNSAGKQKSGKASVATATQAFQESGNGNOTSS
PGIPSYNQATKSAKRLQLQSSSPRLQGTTEKASRRYSLPSSGNSAKITSHSPKTRV
SNSGSKGKNTKTLSSREGNGKATPEVKNR"
complement (35250..40561)
/gene="F6A14.8"
complement (join(35250..35476,36230..36440,36528..36662,
36742..36803,36897..37110,37199..37384,37491..37583,
37674..38073,38169..38296,38844..38987,39095..39217,
39414..39513,39610..39646,39743..39839,39921..40037,
40123..40177,40266..40561))
/gene="F6A14.8"
/notes="Similar to WEB1/SEC31-like protein transport
protein; Similar to WEB1/SEC31 protein transport protein"
/codon_start=1

```

```

/protein_id="AAF27099.1"
/db_xref="GI:6730704"
/translaton="MDCIKISGRSAFVAIPESFFIAAGTMAGAVDLFSSSANLEIF
Query Match 8.8%; Score 451.8; DB 8; Length 108767;
Best Local Similarity 81.9%; Pred No. 1.7e-61;
Matches 533; Conservative 0; Mismatches 117; Indels 1; Gaps 1;
QY 1 TTTTTCGCGGAAAAATCTCGGGTTTACGTTTTTGGCGGAAAAATCTCGTGTTCACGTTTT 60
DB TTTTTCGCGGAAAAATCAACGGTTTACGTTTTTGGCGGAAAAATCAACGTTTTTACGTTTT 74590
QY 61 TGGCGGAAAAATCTCGGGTTTACGTTTTTGGCGGAAAAATCAACGGTTTACGTTTTTGG 120
DB TGGTGGAAAAATCAACGGTTTACGTTTTTGGCGGAAAAATCAACGAGTTTATATTTTTTGG 74810
QY 121 CGGAAAAATCAACGGTTTATGTTTTTGGTGGAAAAATCAACGAGTTTACGTTTTTCTCAAT 180
DB TGGAAAAATCAACGGTTTATGTTTTTGGTGGAAAAATCAACGAGTTTACGTTTTTCTCAAT 74870
QY 181 TCCATTGCTTGTATATTTAAGAAATTTGAAAAATATTAATTTTAAATTCGTTTAGA 240
DB TCCATCGCTTGTATGTTTAAAGAAATTTGAAAAATATTAATTTTAAATTCGTTTAGA 74930
QY 241 TGTGTTGGTTAAACCTAAATTCGCATTTGGTTTACGATTTTACGTTTGGTTTATTCATTT 300
DB TGTGTTAGTTTAAACTTAAATTCGCATTTGGTTTACGATTTTACGTTTGGTTTATTCATTT 74990
QY 301 TACAAAAATTTGATGGGTAAATTCGATAAACCATGAAACCCATTAACCATTAACCACTAAC 360
DB TACAAAAATTTGATGGGTAAATTCGATAAACCATGAAACCCATTAACCATTAACCACTAAC 75050
QY 361 TCATTTTACTCATCAAAACCAATTCGATCATCACTCATTTTGGTCACTCACTCACTCATTTGAG 420
DB TTATTTTATTCATCAAAACCAATTCGATCATCACTCATTTTGGTCACTCACTCACTCATTTGAG 75110
QY 421 TCAAAAATTTCACTCATTTAGGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 480
DB TCAAAAATTTCACTCATTTAGGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 75170
QY 481 TTGACCCATTTTGACACCCCTCATATGATCATAGTAAATTAATCAATAATCAATAATCAATAAT 540
DB TTGACCCATTTTGATACCCCTTAATTCGATGATGATGATGATGATGATGATGATGATGATGAT 75230
QY 541 TAACTTTTAAACGAAATGTTTTGTAAGTTTCATTTACGTTTATTTATATAAAAAAAT 600
DB CTGAGGATACATGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 75290
QY 601 -ATGCAAAAAGTACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 650
DB TATTCACAAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 75341
RESULT 12
AC027135 80139 bp DNA linear PLN 19-JAN-2001
LOCUS Arabidopsis thaliana chromosome 1 BAC T8E3 genomic sequence,
DEFINITION complete sequence.
AC027135
AC027135.6 GI:12322530
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 80139)
AUTHORS Lin.X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC T8E3 genomic sequence
Unpublished
REFERENCE 2 (bases 1 to 80139)

```

AUTHORS Lin,X. and Kaul,S.
 TITLE Direct Submission
 JOURNAL Submitted (28-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
 REFERENCE 3 (bases 1 to 80139)
 AUTHORS Town,C.D. and Kaul,S.
 TITLE Direct Submission
 JOURNAL Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
 COMMENT On Jan 19, 2001 this sequence version replaced gi:12280901.
 Address all correspondence to:at@tigr.org

BAC clone T8E3 is from Arabidopsis thaliana chromosome 1
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.

Genes were identified by a combination of several methods: Gene
 prediction programs including Genscan+ (Chris Burge,
<http://CCR-081.mit.edu/GNSCAN.html>), GeneMarkHMM (Mark Borodovsky,
<http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant
 of GlimmerM, see Mihaela Pertea,
<http://www.tigr.org/softlab/glimmerm.htm>), and
 GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
 peptide database and the plant EST database at TIGR
<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to
 indicate the level of evidence for their annotation. Genes with
 similarity to other proteins are named after the database hits.
 Genes without significant peptide similarity but with EST
 similarity are named as unknown proteins. Genes without protein
 or EST similarity, that are predicted by more than two gene
 prediction programs over most of their length are annotated as
 hypothetical proteins. Genes encoding tRNAs are annotated by
 tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
 Simple repeats are identified by repeatmasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).
 Location/Qualifiers
 1..80139
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cui_tivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="1"
 /clone="T8E3"
 complement(894..1959)
 /gene="T8E3.14"
 complement(join(<894..1112,1254..1457,1801..>1959))
 /gene="T8E3.14"
 complement(join(894..1112,1254..1457,1801..1959))
 /gene="T8E3.14"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAG51277.1"
 /db_xref="GI:12322548"
 /translation="MAHIMMKVLTNTNDKAKRRLTFLNTRKQGERQDTNASW
 LSDRTIKPLPCSGFRSLPLNELLDVGMFGVNGEVIKVAEVLGVVRKSDVLVE
 TSLVMSIDVNGFQVLPSQVESKSLFEKHPDIASFRKPNHLRTISNLSLSTEI
 LHCQSPETLSIDLANASTLCLIKAGFKLD"
 complement(3047..3695)
 /gene="T8E3.3"
 /notes="predicted by genscan"
 complement(join(<3047..3397,3519..>3695))
 /gene="T8E3.3"
 complement(join(3047..3397,3519..3695))
 /gene="T8E3.3"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAG51268.1"
 /db_xref="GI:12322539"
 /translation="MPLTKLLDKNGFMWNGSVKIVAEGVLGVVGVKSDVLBEETSLV
 NESMYNGFQVLPSQVESKSLFEKHPDIESKLHLKPNHRTYLSNLSLTLEISQS
 PENISNDLANASTLYVTKAGFKLDLWLEKLDKVGETRIGQIEBELDKIQKQVEM
 KHKCVDMALLLEFL"
 complement(5787..6988)
 /gene

gene
 complement(9690..11941)
 /gene="T8E3.6"
 complement(join(<9690..10133,10250..10411,10547..10729,
 10818..10958,11116..11364,11460..11599,11680..>11941))
 /gene="T8E3.6"
 complement(join(9690..10133,10250..10411,10547..10729,
 10818..10958,11116..11364,11460..11599,11680..11941))
 /gene="T8E3.6"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAG51275.1"
 /db_xref="GI:12322546"
 /translation="NATAAIIISVKPPASFCGNNRRDGGGITIARVSRDERPHERIV
 GLSATVLFGLSLRVCSPASARILPPPIVVTEINTEITVAGEENLNEVDEEKL
 ADPEAYKSVYSLTVPKLVALRGVSPFSPWIKFMSQSKRVLRKTKQFNGLEETFFD
 LSKPSKNGKGAATAADMTSIGDSWLSFAIKELIEPMKGIEDODWYKGLSDKWK
 QILRRNYAGEKAPDGTWAVFYRWGTWVIAYKSKFQNYKLAPIEDWADLRPELAG
 RAMVNSPREVVAVLKYMRASNTTDLDSQVGGRLAVEKNLAKMQIRLPDSNNY
 LKAFNVGDVWTVGVKSSDVIIPVAKRMSNTVIIPKSGATLWADLWAIIPVSDGKAE
 OGRGVRGSPILNQWIEFCLOPARSLPFTREVIIPGASPSALDGLPVTPEKTKDRT
 KLDTNLVGPDPPEILSKCEPLEPPEATLSRYLLIEVTVRKQSPRGVLEKLDIVS
 IKVRGFRAKLDSMMKNI"
 12959..16475
 /gene="T8E3.2"
 /notes="contains Pfam profile: PF00069: Eukaryotic protein
 kinase domain"
 join(<12959..13055,13146..13281,13357..13428,13506..13577,
 13677..13748,14043..14114,14198..14269,14497..14585,
 14679..14837,14944..15276,15355..15600,15677..15810,
 15894..16475)
 /gene="T8E3.2"
 join(12959..13055,13146..13281,13357..13428,13506..13577,
 13677..13748,14043..14114,14198..14269,14497..14585,
 14679..14837,14944..15276,15355..15600,15677..15810,
 15894..16112)
 /gene="T8E3.2"
 /codon_start=1
 /product="protein kinase, putative"

/gene="T8E3.4"
 /notes="predicted by genemark hmm"
 complement(join(<5787..6092,6218..6444,6618..6784,
 6882..>6988))
 /gene="T8E3.4"
 complement(join(5787..6092,6218..6444,6618..6784,
 6882..6988))
 /gene="T8E3.4"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAG51271.1"
 /db_xref="GI:12322542"
 /translation="MEDQVEKKITWIKNPSFVQSQAIQSDIFVVGDSKWHLLVAPKG
 NERSTNKLCLSLNVADFQSLNGWRHRIKIRLTVVNMSEKLSQEVQIGFYKKNPH
 IGFQTMPLSKLLDKNGFLVNGDVKIYFVGVLEVGVKSDVLEETLLVHESIDNG
 FOVLPSQVESVNNLFEKHPDIYSEFRSKNPLMRTYLLNDLLCLTEILCSSEBLSTGD
 MANAYSTLCLKAGFKLDLWLEKLEKVCARQVEIDEWKDLTLKKEP"
 complement(7821..9077)
 /gene="T8E3.21"
 complement(join(<7821..8165,8304..8521,8706..8872,
 8971..>9077))
 /gene="T8E3.21"
 complement(join(7821..8165,8304..8521,8706..8872,
 8971..9077))
 /gene="T8E3.21"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAG51272.1"
 /db_xref="GI:12322543"
 /translation="MADQVEKKITWIKNPSFIQSHAIYFDIFVVGDTKWHLLVAPKG
 YGDSLNKCLSLFLGVDPDDLPSCGWRHLLIYLTVMNSEKLSQEVARGGFYPRSL
 TFGSQMLPTELYGFLVGQVKIYVAEVLGVKSDVLEETLLVHESIDNG
 LPSQVESVNNLFRNHPDIASFRLENLHRTYLLNSLCLTEILCSKPLNGVDLAN
 AHCTTCVTCKAGFKLDLWLEKLEKVEGTRMQLEQNKLKRESLSDDEDLRSV
 KT"
 complement(9690..11941)
 /gene="T8E3.6"
 complement(join(<9690..10133,10250..10411,10547..10729,
 10818..10958,11116..11364,11460..11599,11680..>11941))
 /gene="T8E3.6"
 complement(join(9690..10133,10250..10411,10547..10729,
 10818..10958,11116..11364,11460..11599,11680..11941))
 /gene="T8E3.6"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAG51275.1"
 /db_xref="GI:12322546"
 /translation="NATAAIIISVKPPASFCGNNRRDGGGITIARVSRDERPHERIV
 GLSATVLFGLSLRVCSPASARILPPPIVVTEINTEITVAGEENLNEVDEEKL
 ADPEAYKSVYSLTVPKLVALRGVSPFSPWIKFMSQSKRVLRKTKQFNGLEETFFD
 LSKPSKNGKGAATAADMTSIGDSWLSFAIKELIEPMKGIEDODWYKGLSDKWK
 QILRRNYAGEKAPDGTWAVFYRWGTWVIAYKSKFQNYKLAPIEDWADLRPELAG
 RAMVNSPREVVAVLKYMRASNTTDLDSQVGGRLAVEKNLAKMQIRLPDSNNY
 LKAFNVGDVWTVGVKSSDVIIPVAKRMSNTVIIPKSGATLWADLWAIIPVSDGKAE
 OGRGVRGSPILNQWIEFCLOPARSLPFTREVIIPGASPSALDGLPVTPEKTKDRT
 KLDTNLVGPDPPEILSKCEPLEPPEATLSRYLLIEVTVRKQSPRGVLEKLDIVS
 IKVRGFRAKLDSMMKNI"
 12959..16475
 /gene="T8E3.2"
 /notes="contains Pfam profile: PF00069: Eukaryotic protein
 kinase domain"
 join(<12959..13055,13146..13281,13357..13428,13506..13577,
 13677..13748,14043..14114,14198..14269,14497..14585,
 14679..14837,14944..15276,15355..15600,15677..15810,
 15894..16475)
 /gene="T8E3.2"
 join(12959..13055,13146..13281,13357..13428,13506..13577,
 13677..13748,14043..14114,14198..14269,14497..14585,
 14679..14837,14944..15276,15355..15600,15677..15810,
 15894..16112)
 /gene="T8E3.2"
 /codon_start=1
 /product="protein kinase, putative"

/protein_id="AAG51266.1"
/db_xref="GI:12322537"
/translation="MGICEMKSCCSWLLILSLCSLNSQASIPDGEALLSPFNNAV
RSDSFIRQWRPDPCCNNGWVTCDAKTRVITLNLTYHKIMGLPDPD1GKLDHLRL
MLHNALYGAIPALGNCTALEEHLQSNFTGPIPAEMGDLPLGLQKLDVSNVTLSP
IPASLGLQKLLSNFVNNTFLVGGIPDQVLSVFSKNSFTGNLNCQKHDVVVQDDSD
GNPSHSGSQGNQKNGSKLLISASATVAGALLIIVMCFWGCFLYKLGKVEIKSLAK
DVGGASIVMFHGDPLPYSSKDIKKLEWNEEHLIGCGFGTYVKLAMDQKVPALKR
ILKNKGFDPPEPPELILGSIKRYILVNLRGYCNSTPSKLLYDIPGSLDEALKE
RGQLDNDRVNIIIGAKGLSYLHDCSPRIIHRD1KSSNLLDGNLEARVSDFGHA
KLEDESHITTVAGTGYLAPEYMQSGRATEKTDVYFGLVLEVLSGRPTDASF
IEKGLNVGMLKFLISEKRPDIIVDPCNCEGMQESLDALLSIATQCVSPSPERTMH
RVQVLESEYMTPCPSFEYDSSD"
/complement(16532..18333)
/gene="T8E3.8"
/note="contains multiple PPR domains: PF01535: PPR repeat"
complement(<16532..>18333)
/gene="T8E3.8"
complement(16621..18333)
/gene="T8E3.8"
/codon_start=1
/product="ppr-repeat protein"
/protein_id="AAG51260.1"
/db_xref="GI:12322531"
/translation="MNNSLIQTPSLMNMNKLKSLADKSFVKLALFGELGGGLYP
DNFTLPVLKISIGRLKVRIGKGVHGVAVKAGLEFDSVNSLMGMVYASLGKIEITHK
VFEMPDQVDSVNGLSVYVNGRPEDALGVKPKMSQESNLKFDGTVLSTLSASA
LKNLEIGERYFVTFEPMSVRIGNALDMFCGCLDKARAVFDSMRDKNVKWTSS
MVFVGTSGIDBARVLFSPVKDVLVTAMNGVYQFNRFDEALEFLFCMGTAGIR
PNFLVLSLTCAQTGALSOQKWIHGYINENRTVDKVVGTALVDMYAKCGCIEIAL
EVPYERKERTASWTSLIYGLMNGMSGRALDYEMENVRDLDAITFVAVLTACNH
GGFVABERKTFHSMTERHNVPKSEHCSCLIDLLCRAGLDEAEELIDKWRGSDTEL
VPVYLSLSAARVNGVNAERVAEKLEKVEVSDSSATILLASVYASANKWEDVTVNR
RKMKDGIKRFPCSSSIDGVGHEFIVGDLLSHPNWDEINSLHQTNNMLDLSEHK
EIDS"
complement(18475..21274)
Query Match
Best Local Similarity 8.7%; Score 449; DB 8; Length 80139;
Matches 491; Conservative 0; Mismatches 30; Indels 4; Gaps 2;
QY 1 TTTTGGCGGAAATCTCGGTTTACGTTTTTGGCGGAAATCTCGTTTACGTTTT 60
Db 49768 TTTTGGCGGAAATATATGATTTACATTTTGGCGGAAATCTCGGTTTACGTTTT 49827
QY 61 TGGCGGAAATCTCGGTTTACGTTTTTGGCGGAAATCTCGGTTTACGTTTTT 120
Db 49828 TGGCGGAAATCTCGGTTTACGTTTTTGGCGGAAATCTCGGTTTACGTTTTT 49887
QY 121 CGGGAATAATCAGGATTTATGTTTTTGGTGGGAAATACGAGTTTACTTTTCTCAAT 180
Db 49888 CGGGAATAATCAGGATTTATG--TTTGTGGTGGGAAATACGAGTTTACTTTTCTCAAT 49946
QY 181 TCAATTCCTGTATATTAGAAATTTGAAATAATTAATTTAATAATGTTTAGA 240
Db 49947 TCAATTCCTGTATATTAGAAATTTGAAATAATTAATTTAATAATGTTTAGA 50006
QY 241 TGTGTTGTTTAACTTAATTTGAGTTTACGTTTACGTTTACGTTTACGTTT 300
Db 50007 TGTGTTGTTTAACTTAAATTTGAGTTTACGTTTACGTTTACGTTTACGTTT 50066
QY 301 TACAAAATTTGATGGGTTAATTTGATAAACCATTGAAACCATTAACACCTTAC 360
Db 50067 TACAAAATTTGATGGGTTAATTTGATAAACCATTGAAACCATTAATTAACACCTTAC 50126
QY 361 TCATTTTACTCATCAACCAATTTGACTCAATCACTCATTTGACTCATCACTCATTTGAG 420
Db 50127 TCATTTTACTCATCAACCAATTTGAC---TCAACTCATTTGACCACTCACTCATTTAG 50183
QY 421 TCAAAAATTTCACTCATTTAGGTTTACGTTTACGTTTACGTTTACGTTTACGTTT 480
Db 50184 TCAAAAATTTTCAATTCATTAGGTTTACGTTTACGTTTACGTTTACGTTTACGTTT 50243

QY 481 TTGACCCATTTTGACACCCCTACATATGATCATTAAGTAAATAATC 525
Db 50244 TTGACCCATTTTGACACCCCTACATATGATCATTAAGTAAATAATC 50288
RESULT 13
AC006551
LOCUS Arabidopsis thaliana chromosome I BAC F12K8 genomic sequence,
complete sequence.
ACCESSION AC006551
VERSION AC006551.6 GI:6056182
KEYWORDS HTG
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 107200)
Fedoritskiy,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Unpublished
2 (bases 1 to 107200)
Fedoritskiy,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (10-FEB-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
3 (bases 1 to 107200)
Fedoritskiy,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (16-OCT-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
4 (bases 1 to 107200)
Fedoritskiy,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altati,H., Araujo,R., Huizar,L., Rowley,D., Brooks,S., Buehler,E.,
Chao,Q., Dunn,P., Gonzalez,A., Khan,S., Kremenetskaia,I., Kim,C.,
Lenz,C., Li,J., Liu,S., Luros,S., Schwartz,J., Shinn,P.,
Toriumi,M., Vyotskaia,V., Yu,G., Ecker,J., Theologis,A. and
Davis,R.W.
Direct Submission
Submitted (16-DEC-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Oct 16, 1999 this sequence version replaced gi:5881526.
Bases 1-12905 of clone F12K8 overlap with bases 70259-83163 of TAMU
BAC clone T22J18, gb|AC003979.
FEATURES
Location/Qualifiers
1..107200
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="I"
/clone="F12K8"
complement(816..1713)
/gene="F12K8.1"
complement(join(816..1326,1407..1536,1614..1713))
/note="Location of ests 250F1477 (gb|AA598046), 250F1577
(gb|AA598042), 110A1477 (gb|T42000.5263), and 184T7P
(T04118). 58% identical to DNA-binding protein


```
/product="hypothetical protein"
/protein_id="AAG60155.1"
/db_xref="GI:12597845"
/translations="MAGISVGLGFFRTLSKAATVAATVAGRGKGIKTRPVSQOLA
SFAGRELRITGSALKKWEYVKLHNLQNPAKKEIHCDCLKITFDCKDKVGIETMK
LLSPHPKSV"
gene
7467. .14213
/genes="F27M3_5"
/notes="contains similarity to cytochrome c"
join(<7467. .7707,8400. .>8689)
mRNA
/genes="F27M3_5"
join(7467. .7707,8400. .>8689)
CDS
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG60137.1"
/db_xref="GI:12597827"
/translations="MGKDDHEDQKDFSHNHGHGYPGPAGPPPPQAGYPPPGVYP
PGYPPPPHGYPPAAVPPPGAYPGSPGPRPGFGGGVGLIAGATAAAAMG
GHHAGHHGGYHGHGKYGKRGFFGGGKYKGGKSMFGGGKYKRGKMGFGGKSKHGM
PGKRGKGMGRKWK"
gene
join(<9659. .9811,9913. .10121,10869. .10935,11021. .11113,
11276. .11419,11508. .11596,11758. .11863,11957. .12044,
12129. .12247,12334. .12495,12603. .12903,12990. .13099,
13178. .13358,13493. .13622,13826. .>14213)
/genes="F27M3_5"
CDS
join(9659. .9811,9913. .10121,10869. .10935,11021. .11113,
11276. .11419,11508. .11596,11758. .11863,11957. .12044,
12129. .12247,12334. .12495,12603. .12903,12990. .13099,
13178. .13358,13493. .13622,13826. .14213)
/genes="F27M3_5"
/protein_id="AAG60136.1"
/db_xref="GI:12597826"
/translations="MVSLSFILCVLVSCAVATIVSHDGRAITIDGHRRLVLSGSIH
YPRSTTEMPDLIKKKGSLDAITVFWNAHEPRROVDPSGNDLIRFLKTIQNE
GMVGLRIGPVCAEWYNGVPPVLMNMGMBERTNTAFMNEQNPMTMIVEMVKE
KLPVAGGGPILIAQIENEGNVIGSGEAGKAIQCANWANSLDVGVFWIMCQDDA
QPMLNTNGYICDNFSNNPNTKWNTEWGNWGGDKPHRTEDVAFVAFPF
QKGGTQNYMYHGFNFRTAGPYITTYDYDAFLDEFNLNPKYGHKLQHDVL
HAMEKRTIYGNISTVDFGLVATVYQTEBEGSCFIGNVNETSDAKINFQTSYDPA
WSVSLDPCKTEYNTAKNTQTSVMVKANAEANBPSTLKWSMRPENTIDSVLLKGGK
ESTMROLFDQKYNDESXYLWMTTVNLKBPQVLGKNLSLRINSTAHVLFVNGQ
HIGNYVENGKEHVVEQDAKPGANVAILLSITVGLNYGAFNFENRAGITGPVFI
IGRNGDETIVKDLSTKWSYKTLGSLFENQLFSSEPSWNSAPLSEPVVDLLGK
GTAWINGNIGVYAPFLSDIDGCSAEYHVPSFLNSEGNDLVLFEETGGNPSLVNF
QITGVGSVCANVYKKNVLESLCNGKPISAIKFASFGNPGDGCDFEKGTCESNNAAA
IITQECVKEKCSIDVSEKFGAEGCALAKRLAVEAIC"
complement(15582. .19591)
/genes="F27M3_7"
/notes="similar to epsilon adaptin GI:5689377 from (Homo
sapiens)"
complement(join(<15582. .16556,16628. .16771,16854. .16931,
17301. .17417,17501. .17557,17793. .17924,18017. .18214,
18313. .18423,18506. .18706,18803. .>19591))
/genes="F27M3_7"
CDS
complement(join(15582. .16556,16628. .16771,16854. .16931,
17301. .17417,17501. .17557,17793. .17924,18017. .18214,
18313. .18423,18506. .18706,18803. .19591))
/genes="F27M3_7"
/codon_start=1
/product="epsilon-adaptin, putative"
/protein_id="AAG60138.1"
/db_xref="GI:12597828"
/translations="MSQLKTIHGLAMGQGGFGQSKFELDLVKSICEARSKAEDRI
VLSEVILLKRLLEPDKRMKEYILRLVIEMLGHDASFGYIYAVKMTDDNLLLK
RTGLYATVFLNEDHDLILIVNTIOKDLSDNLYVCAALNAICELINEERTPAVLK
OVVELLNHOKAEVRKKAIMALHPRKSPSSVSHVSNFRKLCNDPVGWAGATLCP
FDLISEVNSYKDLVSVSLKQVTERLPKSIDYHQMPAPFIQIMALLGSGDKNAS
DIMSIVGLDFRCKDSSTIGNAILYECIRCSILPNPKLLEAAADAIKFLKDSH
NLKXMGIDGLKIKISPDIAEQHQLAVIDCLEDDPDLTKRKTFFELLYKMTKSNVEV
IVDRMIDYIMISINDNHYKTEIASRCVELAEQAPAPSNQWFIQIMNKVFEHAGDLVNIK
```

```
ANMLRLIAGEGEDDDADSKLRLSAVESYLIQISEPKLPSLFIQVSWLGEYGTGTA
DKYISYISGLKCDVADAYSDETVKGYAVSALMKIYAFIASGRKVDVLPQCQLI
ELLASHSDLOQRAYELQALLADAVETIPLDASCEIEVDKDLGFLNGYIQQA
IFSGAPYISERESGMFETDYPQDHEVTHALRFAYELPKFESVPPQASNEIIVP
VPEPSYSSHQPISTLSVRESSEBIKLRLDGVKOKWRPSYVSTTAASSTTPOAAN
GLTSHSDAGVSSSSKPRSSYRPPKPEIDPEKORLAASLFGGSSSTDKRSSGGHKP
AKGTANKTATVPEKNTQTPVQPPDLLDFGEPTATTATAMDPEKLEGLMDSSQDGS
SDVMGLYSQAPVPTTISVLSLSLSESSKSNRTYQPQTSKGPNTKEALEKDALVR
QMGVNPTSDRPTLFDKLLG"
complement(22142. .23329)
/genes="F27M3_8"
/notes="predicted by genemark hmh"
complement(join(<22142. .22194,22558. .22775,22906. .23126,
23204. .>23329))
/genes="F27M3_8"
CDS
complement(join(22142. .22194,22558. .22775,22906. .23126,
23204. .23329))
/genes="F27M3_8"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG60140.1"
/db_xref="GI:12597830"
/translations="VEIQODNRDGRPKSFLPFMFIFLFLGLAAFFLCLSAEFOKAKAL
LRAOVFLKGLDKWDCECYLPENRAFGIGIALVCVSAQIVGNVIVIRGFTKDTKT
RTIIFCIILLPSWNVFAVATLISVGSAMREQIVKGWLNRECYLVKDGVPAAAGF"
Query Match 8.7%; Score 449; DB 8; Length 110664;
Best Local Similarity 93.5%; Pred. No. 4.7e-61;
Matches 491; Conservative 0; Mismatches 30; Indels 4; Gaps 2;
QY 1 TTTTTCGGCGGGAATACTCGGGTTTACGTTTTCGGCGGGAATACTCGGGTTTACGTTT 60
Db 92326 TTTTTCGGCGGGAATACTCGGGTTTACGTTTTCGGCGGGAATACTCGGGTTTACGTTT 92267
QY 61 TGCGCGGGAATACTCGGGTTTACGTTTTCGGCGGGAATACTCGGGTTTACGTTTTCG 120
Db 92266 TGCGCGGGAATACTCGGGTTTACGTTTTCGGCGGGAATACTCGGGTTTACGTTTTCG 92207
QY 121 CGGGAATACTACGGAATTAATGTTTTCGGCGGGAATACTACGGAATTAATGTTTTCG 180
Db 92206 CGGGAATACTACGGAATTAATGTTTTCGGCGGGAATACTACGGAATTAATGTTTTCG 92148
QY 181 TTCATTGCTTGTATATTTAAGAAATTTGGAATAATTAATTTAATAATTTGTTTAGA 240
Db 92147 TTCATTGCTTGTATATTTAAGAAATTTGGAATAATTAATTTAATAATTTGTTTAGA 92088
QY 241 TGTGTTGGTTAAACCTAAATTCGTCATTGTTTAGAGATTTCGTTTTCGTTTTCGTTTTC 300
Db 92087 TGTGTTGGTTAAACCTAAATTCGTCATTGTTTAGAGATTTCGTTTTCGTTTTCGTTTTC 92028
QY 301 TACAAAATTTGATGGGTTAATTCGATAAACCATGGAACCATTAACCATTAACCATTAAC 360
Db 92027 TACAAAATTTGATGGGTTAATTCGATAAACCATGGAACCATTAACCATTAACCATTAAC 91968
QY 361 TCATTTTACTCATCAACCAATTCGATCATCACTCACTCACTCACTCACTCACTCACTCACTT 420
Db 91967 TCATTTTACTCATCAACCAATTCGATCATCACTCACTCACTCACTCACTCACTCACTT 91911
QY 421 TCAAAAATTTCAACTCATTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 480
Db 91910 TCAAAAATTTCAACTCATTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 91851
QY 481 TTGACCCATTTTGACACCCCTACATATAGTATAGTATAGTATAGTATAGTATAGTATAGT 525
Db 91850 TTGACCCATTTTGACACCCCTACATATAGTATAGTATAGTATAGTATAGTATAGTATAG 91806
RESULT 15
AC022522/c
LOCUS
DEFINITION
Arabidopsis thaliana chromosome I BAC T28K15 genomic sequence,
complete sequence.
AC022522
AC022522
```


CDS	complement (24845..27514) /gene="T28K15.5" /note="NBS/LRR disease resistance gene, location of genomic clone gb AF074916." /codon_start=1 /product="resistance to Pseudomonas syringae protein 5" /protein_id="AG12572.1" /db_xref="GI:10086512" /translation="MGGCFSVSLPCDDVQVVSQQLCVRGSVYHNLNKLASLQKAMR MLKARQYVIRLEETFEFTGRQRLSQVQVLTSLIIQNFNDLLRSNEVELRLCL CGFCGKDLVLYKRVKRWMLKEVESLSQGFDFVVSSEATPFDVDEIPFOPTVGO ETLMLKAWNLMEKDSGIIAGLXMGVGKTTLLTKINNFKSKIDRFEDVVIWVVSRS STVKTQRLAKVGLGGMSEKDNQIADVHVLRRKFEVLDDLDLWEKVNKLAV GPYFSKQNGKFAFTIRSDVCGRMGVDDPNEVSCLPQESWDLPQMKVGNKTLSGH PDIPGLARVAKRCRGLPALNVIGEAMACKRTVHEWCHADIVLTSADFSGMDEIL LEVLKYDNLNGLMKSFCLYCSLPEDYLLDKGLVDYWI SEGFINEKEGRERNIN QYGEYIITVLRAKLEEBERNKSNVQKHDVREMAWISSDLGKQKEKICVIRAGVGLR EYVKVDWNTVAKI SLNMNEI BEI FDSHECAATLTLFLQKNDVVKI SABFFPCMPHLV VLDLGENOSLNEIPEIISLASLRVENLYCTIHLPGVGLTKLHLNLEHMSLIG SILGNSNLNLTGLRDSRLLDMSLVLEQLLHLEVLITDIDISSLVAEPLLSQR LVECHKEVDKYLKEESVRLTLPNTGNLRKLGKRCGMREIKIBRTSSSSNKSPT TPCFNLSRVFTAKCHGLKDLTWLLFAPNLTFLEVFSKEVEDII SEEKAEEHSATIV PRKLETLHLFRLGKRIYAKALHPCLKVIHVKECKELRKLPLDLSKSGIAEEIIVI YGEREWIERVWEDQATOLRFLPSSRWRET" complement (28920..31577) /gene="T28K15.6" complement (28920..31577) /gene="T28K15.6" /note="Similar to RPS5, location of genomic clone gb AF074916." /codon_start=1 /product="NBS/LRR disease resistance protein" /protein_id="AG12573.1" /db_xref="GI:10086513" /translation="MGGCVSVSLSCDREVNQFSQMLCVSGSVYQNLSENLSLQKAMG VLNAKRDVQGRINREEFTGHRRLAQOVQWLTIQTIENQFNLDLSTCNABIQRLCL CGFCNSKNVMSVLYKRVILVAREVGLSSQGVDFIVTEAAPAEVEBELPIOSTIVGQ DSNLDKVNCLMEDKWIIVGLYMGVGKTTLLTQINNFKESKLGGFDVVIWVVSQN AVTHKIQGIGKGLVGRNWDENKQNALDIHNVLRERKFLVLLDDDIWEKVELKVI PVIQARVSKKCCGLPALNVIGTMSFKRTIQEWHRATEVLTSATDFSGMEDEIL PILKYSYDSLNGEDAKSCFLYSLFPDEFIRKEMLI EYWICBGFIRKEKQREKAFNQ GYDILGTVRSSLLLGAKDKOVSMHDMVREMAWIFSDLGKHKERCIVQAGIGLDE LPEVNRVAKRSLMNNNFELKSPFCVELLITFLQNNYKLVDSIMBFFRCMPSLA VLDLSNLSLSEIPEISELVLSQYLDLSGLTYIERLPHGLHRLKLVHLKLETRLE SLSGYSYLSRLRDLRDEFTLDGIMKELQLLEHLITDIDISSGLVGEFLFCYPRV GRICHIYIRDPWERPEESVGLVLPALHNLCYISINWCMMWMIIEKTPFWKNLTNP NFSNLGNVRTEGCDGLKDLTWLLFAPNLNLRVWGCKHLEDIISKEKAASVLEKETLP FOKLECLNLYQLSELKSIYWNALPFORLRCLDILNCPKLRKLPLDLSKSVVKEEPIV KYEKKWIERVWEDEATQYRFLPTCLR"	Db	78409	ATGTGTTGGTTAAACTTAAATTTGACATTTGGTTTAGAGATTTTAGTTGGTTTAATCAATT	78350
		QY	300	TTACAAAAATTTGATCGGTTAAATTTGGATAAACCAATGGAACCAATTAACCAATTAACCACTTAA	359
		Db	78349	TTACAAAACTTATTGGGTTAAATTTGGGTAACCAATTAACCAATTAACCAATTAACCACTTAA	78290
		QY	360	CTCATTTTTACTCATCAACCAATTTGACTCATCACTCATTTGACTCATCACTCATTTTGA	419
		Db	78289	CTCATTTTACTCATCAACCAATTTGACTCATCACTCATTTGACCACTCATTTGACCACTCATTTGA	78230
		QY	420	GTCAAAAAATTTCAACTTCATTTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTG	479
		Db	78229	GTCAAAAAATTTTAACTCATTTAGGGTTTCATGGATTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTG	78171
		QY	480	TTTGAACCAATTTTGAACCCCTACATATGATCATTAAGTTAATAACAAAAATTAATTAATTAATTAATTA	539
		Db	78170	TTTGAACCAATTTTGAACCCCTAGTTAGTTATAGAAACCTGTAACCACTGTAACCACTGTAACCACTG	78111
		QY	540	ATAAATTTTAAACGGAATTTGTTTCTGTAAGTTTCAATTTAGTTTATATATATATATATATATATATAT	599
		Db	78110	ATAAATTTTAAATCAACATTTAATATGTTTTCTTAATATGTTGTTTTCTTAATATGTTGTTTTCTTAAT	78051
		QY	600	AATGCAAAAGTACTAATGTATATAGTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT	659
		Db	78050	AAACAAAAATGAACGGGAGGAGTACTAATTAAGTAATTAAGTAATTAAGTAATTAAGTAATTAAGTAAT	77991
		QY	660	ATACTTTTTCATTTCTTATCAATATTTTCTTT	691
		Db	77990	CACCTAGTTAAGCTCATATAAACTAATTTCTCATT	77959

Search completed: October 21, 2003, 09:29:30
Job time : 18167 secs

Query Match	8.7%;	Score 448;	DB 8;	Length 88643;
Best Local Similarity	80.8%;	Pred. No. 7e-61;		
Matches	559;	Conservative	0;	Mismatches 130; Indels 3; Gaps 3;
QY	1	TTTTTGGCGGAAAATCTC-GGGCTTTACGTTTTTTGGCGGAAAATCTCGTGGTTTACGTTT	59	
Db	78648	TTTTTGGCGGAAAATCACAGGATTTACATTTTGGCGGAAAATCTTGGGTTTACGTTT	78589	
QY	60	TTGGCGGAAAATCTCGGGTTTACGTTTTTTTGGGAGAAAATCACGGGTTTACTTTTTTT	119	
Db	78588	TTGGCGGAAAATCTTGGGTTTACGTTTTTTTGGGAGAAAATCACGGGTTTAC-TTTTTTG	78530	
QY	120	GCAGGAAAATCACGGATTTATGTTTTTTGGTGGAAAATTAACGAGTTTACTTTTTCTCAA	179	
Db	78529	GCAGGAAAATCACGGGTTTATGTTTTTTGGTGGAAAATTAACGAGTTTACTTTTTCTCAA	78470	
QY	180	TTTTCATTCCTTGTATATTTAAGAAAATTTGAAAAATTAATTTTATTAATTTGGTTTAC	239	
Db	78469	TTTTCATTCATTCGTATATTTAAGAAAATTTGAAAAATTAATTTTATTAATTTGGTTTAC	78410	
QY	240	ATGTGTTGGTTAAACCTTAAATTTGGCATTTGGTTTGTAGAGATTTTAGTTGGTTTATCAAT	299	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2003, 00:52:07 ; Search time 9739 Seconds

(without alignments)
12837.275 Million cell updates/sec

Title: US-09-824-735-1

Perfect score: 5144

Sequence: 1 tttttgggggaaatctcg.....acgttttgatcccaacttaa 5144

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 4562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estlin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estio.*

8: em_estc.*

9: gb_est1.*

10: gb_est2.*

11: gb_est3.*

12: gb_est4.*

13: gb_est5.*

14: gb_est6.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gssi.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	446.2	8.7	474	29	CNS00R7B
C 2	424.8	8.3	690	28	AO961396
C 3	423	8.2	506	29	AL941773
C 4	375.2	7.3	433	29	CNS00SPA
					AL086741 Arabidops
					AO961396 LERFL01TF
					AL941773 Arabidops
					AL086684 Arabidops

5	346.2	6.7	751	28	AO961397
6	328	6.4	457	28	BH254517
7	320.6	6.2	531	28	AO966842
8	292.4	5.7	363	29	CC055323
9	266	5.2	441	29	CNS00XC3
10	259.8	5.1	319	28	BH902224
11	231.4	4.5	528	28	BH237177
C 11	216.6	4.2	367	29	BX289638
C 12	194.2	3.8	598	28	B77189
C 13	185.8	3.6	514	9	AV830052
C 14	173.6	3.4	454	29	BX291485
C 15	172.6	3.4	204	28	BH910603
C 16	168.8	3.3	819	29	BZ456258
C 17	168.8	3.3	819	29	BZ456258
C 18	150.2	2.9	476	29	CC460177
C 19	150	2.9	816	28	BH599078
C 20	147.2	2.9	302	28	BH757375
C 21	147	2.9	672	28	B20959
C 22	121.8	2.4	133	29	BZ664600
C 23	121.8	2.4	133	29	CC459381
C 24	121.2	2.4	419	29	CC460016
C 25	115.2	2.2	168	28	BH905489
C 26	110.8	2.2	351	29	AL771439
C 27	109.8	2.1	435	29	AL953182
C 28	109.4	2.1	698	28	BH959869
C 29	109	2.1	422	9	AV809816
C 30	108.8	2.1	509	29	CNS00NAV
C 31	108.8	2.1	589	20	BF005416
C 32	107.2	2.1	447	28	BH213915
C 33	106.6	2.1	726	12	B1309472
C 34	105.6	2.1	584	10	BE801990
C 35	105.2	2.0	740	12	BM985643
C 36	104.4	2.0	670	13	BQ587842
C 37	103.6	2.0	1200	13	BX437758
C 38	103.2	2.0	404	14	CA798037
C 39	101.6	2.0	435	29	CC055495
C 40	101.4	2.0	506	13	BQ104557
C 41	99.8	1.9	422	9	AI776493
C 42	99.8	1.9	492	9	AW738389
C 43	99	1.9	1201	9	AL514129
C 44	98	1.9	401	12	BQ606265
C 45	97.8	1.9	255	10	BE248163

ALIGNMENTS

RESULT 1
CNS00R7B/c 474 bp DNA linear GSS 28-JUN-1999
LOCUS Arabidopsis thaliana genome survey sequence SP6 end of BAC F13C22
DEFINITION of IGF library from strain Columbia of Arabidopsis thaliana,
Genomic survey sequence.
ACCESSION AL086741
VERSION AL086741.1 GI:5287881
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 474)
AUTHORS Salanoubat,M., Choisme,N., Artiguenave,F., Brottier,P., Wincker,P.,
Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 474)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
FEATURES
source
1..474
/organism="Arabidopsis thaliana"

```

/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="F13C22"
/clone_lib="IGP"
/notes="end: SP6"

BASE COUNT      163 a   93 c   73 g   145 t
ORIGIN

Query Match      8.7%; Score 446.2; DB 29; Length 474;
Best Local Similarity 98.3%; Pred. No. 9.5e-46;
Matches 470; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 3274 AGTTTAAATACATAGATGACTGACCCCAATCCAAAACAGTGAGTATTTTCCTTGTT 3333
DB 474 AGTTTAAATACATAGATGACTGACCCCAATCCAAAACAGTGAGTATTTTCCTTGTT 415
QY 3334 CTCTCCTAGTATCAGGTTTGGTGATATTAATGTT-CTAGTAAATTAATATCTGTTTATC 3392
DB 414 CTCTCCTAGTATCAGGTTTGGTGATATTAATGTTTCTGGTAATATATCTGTTTATC 3355
QY 3393 TATTATGTTTCTCAATAGACAGCGTATTCAAAATCAAGGAATCAAGAAAGATCCTTG 3452
DB 354 TATTATGTTTCTCAATAGACAGCGTATTCAAAATCAAGGAATCAAGAAAGATCCTTG 295
QY 3453 GTTCAGATTAAATTAATGTCCTATACGACGAGGAAGAGAAGTGAATTTGGATGA 3512
DB 294 GTTCAGATTAAATTAATGTCCTATACGACGAGGAAGAGAAGTGAATTTGGATGA 235
QY 3513 TATTCGTCGAGTTTTCATGGAATGAGGTTTGCTTCGCTTCATTATATGCTCTTTG 3572
DB 234 TATTCGTCGAGTTTTCATGGAATGAGGTTTGCTTCGCTTCATTATATGCTCTTTG 176
QY 3573 CTGTCATATCCATTTAAATGTTAAAGATCTCTTAGGAACGTTTGGATGACCAAGAAG 3632
DB 175 CTGTCATATCCATTTAAATGTTAAAGATCTCTTAGGAACGTTTGGATGACCAAGAAG 116
QY 3633 TGTTGCTACAGATAGAACAAATAGTAGCATCTGTTAGTTACCAAACTGTAACTG 3692
DB 115 TGTTGCTACAGATAGAACAAATAGTAGCATCTGTTAGTTACCAAACTGTAACTG 56
QY 3693 CTTCCTTTATTCATTCGCCAAACCATAGACTTAGGAAGACTTAGATTACAAGA 3747
DB 55 CTTCCTTTATTCATTCGCCAAACCATAGACTTAGGAAGACTTAGATTACAAGA 1

RESULT 2
AQ961396/c
LOCUS      690 bp      DNA      linear      GSS 28-JAN-2000
DEFINITION LRF01TF LREA Arabidopsis thaliana genomic clone LRF01, genomic survey sequence.
ACCESSION  AQ961396
VERSION     AQ961396.1 GI:6789097
KEYWORDS   GSS.
SOURCE     Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
            ; eurosid II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 690)
AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
JOURNAL   Unpublished
COMMENT   Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For additional information, see http://www.tigr.org/cdb/at/at.html

Seq primer: TF
Class: shotgun.
Location/Qualifiers
1..690
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/clone="LRF01"
/clone_lib="LERA"
/notes="Organ: Leaf; Vector: pHO51; Total genomic DNA was sheared to 0.9-1 kbp before ligation."
BASE COUNT      257 a   112 c   88 g   233 t
ORIGIN

Query Match      8.3%; Score 424.8; DB 28; Length 690;
Best Local Similarity 88.2%; Pred. No. 3.7e-43;
Matches 485; Conservative 0; Mismatches 62; Indels 3; Gaps 2;

QY 2 TTTTGGCGGAAAATCTCGGTTTACGTTTTCGGCGGAAAATCTCGTGTTCACGTTTTT 61
DB 678 TTTTGGCGGAAAATCTCGGTTTACGTTTTCGGCGGAAAATCTCGGTTTACGTTTTT 619
QY 62 GCGCGAAAATCTCGGTTTACGTTTTCGGCGGAAAATCAACGGGTTTACTTTTTTTCG 121
DB 618 GCGCGAAAATCTCGGTTTACGTTTTCGGCGGAAAATCAACGGGTTTACTTTTTTTCG 560
QY 122 GCGAAAATCAACGGATTTATGTTTTTTCGGGAAAATTAACGAGTTTACTTTTTCTCAAT 181
DB 559 GCGATAATCAACGGGTTTATGTTTTTTCGGTGGAAAATTAACGAGTTTACTTTTTCTCAAT 500
QY 182 TCATTCCTTGATATTTTAAAGAAATTTGAAAATATTAATTTTAAATTTGTTTTAGAT 241
DB 499 TCATCCCTTGATATTTTAAAGAAATTTGAAAATATTAATTTTAAATTTGTTTTAGAT 440
QY 242 GTCTTGTTTAAACCTAAATTTGCATTTGGTATTTAGAGATTTTAGTTGTTTTTAACTTTT 301
DB 439 GTCTTGTTTAACTTAAATTTGCATTTGGTATTTAGAGATTTTAGTTGTTTTTAACTTTT 380
QY 302 ACAAATTTGATGGTTTAAATTTGATAAACCATGGAACCATTAACCATTAACACCTAACT 361
DB 379 ACAAATTTGATGGTTTAAATTTGGTAAACCATAGAAACCATTAACCATTAACACCAACT 320
QY 362 CATTTTACTCATCAAAACCAATTCATCACTCACTCACTCACTCACTCACTCACTCACT 421
DB 319 CA--TTACTCATCAAAACCAATTCATCACTCACTCACTCACTCACTCACTCACTCACT 262
QY 422 CAAAAATTTCAACTCATTAGGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 481
DB 261 CAAAAATTTCAACTCATTAGGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 202
QY 482 TGACCCATTTTGACACCCCTACATATGATCATAGTTTAAATTAATCAAAAATTAATTTGAT 541
DB 201 TGACCCATTTTGACACCCCTACATATGATCATATGATCATATGATCATATGATCATATGAT 142
QY 542 AACCTTTTAA 551
DB 141 AGCTAAATAA 132

RESULT 3
AL941773
LOCUS      506 bp      DNA      linear      GSS 24-OCT-2002
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-260F08-014949, genomic survey sequence.
ACCESSION  AL941773
VERSION     AL941773.1 GI:24398371
KEYWORDS   GSS.
SOURCE     Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsis.

```

```

REFERENCE
AUTHORS      Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
              and Weissshaar,B.
TITLE        A pipeline for automated high-throughput generation of FSTs
              (flanking sequence tags) from Arabidopsis thaliana T-DNA
              transformed lines
JOURNAL      Unpublished
REFERENCE
AUTHORS      Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weissshaar,B.
TITLE        A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
              for flanking sequence tag based reverse genetics
JOURNAL      Unpublished
REFERENCE
AUTHORS      Strizhov,N., Li,Y., Rosso,M. and Weissshaar,B.
TITLE        Direct Submission
JOURNAL      Submitted (21-OCT-2002) Weissshaar B., Max-Planck-Institut fuer
              Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT      This sequence is recovered from the left border of the T-DNA. It
              indicates an insertion close to or within gene At5G35410. The
              sequences are generated at the MPI for Plant Breeding Research in
              the context of the GABI-Kat project. GABI-Kat is part of the German
              Plant Genomics program designated 'GABI'. Information on line
              availability can be found at:
              http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES     Location/Qualifiers
              source
                1..506
                /organism="Arabidopsis thaliana"
                /mol_type="genomic DNA"
                /strain="Columbia 0"
                /db_xref="taxon:3702"
                /clone="GK-260F08-014949"
                /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                /note="PCR was performed on DNA from Arabidopsis thaliana
                plants (TI) which were transformed with the T-DNA from
                vector pAC161. The lines contain one or more T-DNA
                insertions. The DNA fragment(s) resulting from the PCR
                were directly sequenced to determine the genomic sequence
                flanking the insertion. Sequences displaying significant
                similarity to the A. thaliana nuclear genome sequence were
                processed for submission. T-DNA derived sequences were
                removed"
BASE COUNT   140 a 86 c 103 g 177 t
ORIGIN
Query Match      8.2%; Score 423; DB 29; Length 506;
Best Local Similarity 96.3%; Pred. No. 6.8e-43;
Matches 488; Conservative 0; Mismatches 10; Indels 9; Gaps 5;

QY 3939 TCTTCTAAACTGTCTACTTCAATGTTAACTAACTAACTCAGTATCTGTCTTAGCTAAAG 3998
DB 1 TCTTCAAAAACCTGTACTTAATGTTT---AACTAACTCAGTATCTGTCTTAGCTAAAG 57

QY 3999 TTACTTTTACTGTTTTCATTAAGTTGACGTGCAATGTCACCTTGTTCACAGGCAGTT 4058
DB 58 -TACTTTTACTGTTTAA---TTAGTTGACCTGT-AAATGCACTGTT-ACAGGCAGTT 111

QY 4059 ATCTAGCGGAAGTGTAGAGAAATGATGAGCGGCCCTGATGATGAATGCCCTTCAGA 4118
DB 112 ATGTAGCGGAAGTGTAGAGAAATGATGAGCGGCCCTGATGATGAATGCCCTTCAGA 171

QY 4119 TGATTACTTATCAAGGCTTAAATTTATCTGCACTATTGTCAGCGCAGGTAGTAC 4178
DB 172 TGATTACTTATCAAGGCTTAAATTTATCTGCACTATTGTCAGCGCAGGTAGTAC 231

QY 4179 CTGATTTTCTATTACTGGTCTAGAGATCTCCATTTTGAATPAAAGAAATGCGGTAGCAT 4238
DB 232 CTGATTTTCTATTACTGGTCTAGAGATCTCCATTTTGAATPAAAGAAATGCGGTAGCAT 291

QY 4239 CTATTTCTACAGTCCCGGTTTTCAGTCCCTTATGATGCTGTCTTAGTTTCTTTATAA 4298
DB 292 CTAATTTCTACAGTCCCGGTTTTCAGTCCCTTATGATGCTGTCTTAGTTTCTTTATAA 351

QY 4299 TAACTATAAGTTTCATTAGATGATTGGTTGTCATGGCAATTAGTAGATACAAATGGAATCAA 4358

```

```

DB 352 TAACTATAAGTTTCATTAGATGATTGGTTGTCATGGCAATTAGTAGATACAAATGGAATCAA 4311
QY 4359 AATGTTTCTGCAATATGATGGCTGATCCTTTTGATCTCGCAGGATTTTGTATAAAGCAA 4418
DB 412 AATGTTTCTGCAATATGATGGCTGATCCTTTTGATCTCGCAGGATTTTGTATAAAGCAA 471
QY 4419 CCGGTTTTGTTTCTCGAAGGACCTA 4445
DB 472 CCGGTTTTGTTTCTCGAAGGACCTA 498

RESULT 4
CNS00SPA/c
LOCUS
DEFINITION   CNS00SPA 433 bp DNA linear GSS 28-JUN-1999
              Arabidopsis thaliana genome survey sequence SP6 end of BAC T2H7 of
              TAMU library from strain Columbia of Arabidopsis thaliana, genomic
              survey sequence.
ACCESSION   AL088684
              AL088684.1 GI:5289824
VERSION     GSS.
KEYWORDS    Arabidopsis thaliana (thale cress)
SOURCE      Arabidopsis thaliana
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE   1 (bases 1 to 433)
AUTHORS     Salanoubat,M., Choisme,N., Artiguenave,P., Brottier,P., Wincker,P.,
              Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 433)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
              BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
FEATURES     Location/Qualifiers
              source
                1..433
                /organism="Arabidopsis thaliana"
                /mol_type="genomic DNA"
                /strain="Columbia"
                /db_xref="taxon:3702"
                /clone="T2H7"
                /clone_lib="TAMU"
                /note="end : SP6"
BASE COUNT   162 a 75 c 66 g 130 t
ORIGIN
Query Match      7.3%; Score 375.2; DB 29; Length 433;
Best Local Similarity 93.3%; Pred. No. 5.6e-37;
Matches 403; Conservative 0; Mismatches 28; Indels 1; Gaps 1;

QY 79 TTTACGTTTTTTCGAGAAATCACGGTTTACTTTTTTTTGGGGAATAATCAGGATTT 138
DB 433 TTTACGTTTTTTCGCGGAAATCACGGTTTAC-GTTTTTGGCGGGAATAATCAGCGGTTT 375

QY 139 ATGTTTTTGGTCGAAAAATTGAGTTTACTTTTTCTCAATTTCTATGCTGTATATTT 198
DB 374 ATGTTTTTGGTAGAAAAATTACGAGTTTACTTTTTCTCAATTTCTATGCTGTATATTT 315

QY 199 AAGAAATTTGAAAAATTAATTTTAAATTTAGATGTGTTGGTTAAACCTAA 258
DB 314 AAGAAATTTGAAAAATTAATTTTAAATTTAGATGTGTTGGTTAAAGCTTAA 255

QY 259 ATTGGCATGGTTTAGAGATTTTAGTTGGTTTATTCAATTTTACAAAATTTGATGGTT 318
DB 254 ATTGGCATGGTTTAGAGATTTTACTTGGTTTAAATTTCAATTTTACAAAATTTGATGGTT 195

QY 319 AATTGATAAACCATGGAAACCATTAACCATTAACCACTTAACCTATTTCTATCAAAC 378
DB 194 AATTGGGTAACCATTAACCACTTAACCACTTAACCACTTAACCTATTTGACTCATCAAAC 135

QY 379 CAATTGACTCATCAACTCATTTGACTCATCAACTCATTTGAGTCAAAAATTTTCAACTCAT 438

```



```

QY 4699 AATTGTTATGATTGTTACGTTCTATGCTCAACAGACAGAGCTCGAGGATTATCTTCG 4758
|||||
DB 141 AATTGATATGATTGTTACGTTCTATGCTCAACAGACAGAGCTCGATGATTATCTTCA 200
|||||
QY 4759 ATCAAGCCCGGACAGTTAGCTGTTGTGATAGAGGTAATTTATGCTGTTGTGATTGTAAT 4818
|||||
DB 201 ATCAAGCCCGGACAGTTAGCTGTTGTGATAGAGGTAATTTATGCTGTTGTGATTGTAAT 260
|||||
QY 4819 ATAAGTTTCTTTGCTTCAGTTTAAGGTTATCTAGCAAAATTGAATTAACCTACATGCGAG 4878
|||||
DB 261 ATAAGTTTCTTTGCTTCAGTTTAAGGTTATCTAGCAAAATTGAATTAACCTACATGCGAG 320
|||||
QY 4879 AATTAGAGGTGGCCACATCGCTTTTCATGTTAGAGTGAAGAAAGCTCTGTTGAAACT 4938
|||||
DB 321 AATTAGAGGTGGCCACATCGCTTTTCATGTTAGAGTGAAGAAAGCTCTGTTGAAACT 380
|||||
QY 4939 CTTGAATATCAACAAGGTTTATAAATATATATCAATTAACAATAGTTGCAATCAATTTACTGTG 4998
|||||
DB 381 CTTGAATATCGCAAGGCTTATCA--TACTATCCAATGACAGTAGTTGCTTCATCACTGTC 438
|||||
QY 4999 TTGCGGATAGAGTGATA 5016
|||||
DB 439 GCTGCGTTAGAGTGATA 456
|||||

RESULT 7
AQ966842
LOCUS
DEFINITION
  LERIM53TRB LERG Arabidopsis thaliana genomic clone LERIM53, genomic
  survey sequence.
ACCESSION
  AQ966842
VERSION
  AQ966842.1 GI:6794543
KEYWORDS
  GSS.
SOURCE
  Arabidopsis thaliana (thale cress)
ORGANISM
  Arabidopsis thaliana
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
  ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
  1 (bases 1 to 531)
  Buehl,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
  Feldblyum,T., Liang,F., Cressy,T. and Fraser,C.M.
  Genomic survey sequencing of Landsberg erecta ecotype of
  Arabidopsis thaliana and identification of sequence-based
  polymorphisms
  Unpublished
  Contact: Xiaoying Lin
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: at@tigr.org
  For additional information, see http://www.tigr.org/tdb/at/at.html
  Seq primer: TR
  Class: shotgun.

FEATURES
  source
    1..531
      /organism="Arabidopsis thaliana"
      /mol_type="genomic DNA"
      /strain="Landsberg erecta"
      /db_xref="taxon:3702"
      /clone="LERIM53"
      /clone_lib="LERG"
      /notes="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
      sheared to 0.4-0.7 Kbp before ligation."
BASE COUNT
  147 a 68 c 117 g 199 t
ORIGIN
  Query Match 6.2%; Score 320.6; DB 28; Length 531;
  Best local Similarity 91.0%; Pred. No. 2.9e-30;
  Matches 352; Conservative 0; Mismatches 34; Indels 1; Gaps 1;
  QY 1 TTTTGGCGGAAATCTCGGTTTACGTTTTCGCGGAAATCTCGTCTTACGTTT 60
  |||||

```

```

DB 146 TTTTGGCAGGAAATACACGGTTTACGTTTTCGCGAGAAATATGAATTTACGTTAT 205
|||||
QY 61 TGGCGGAAATCTCGGTTTACGTTTTCGCGAGAAATACACGGTTTACGTTTTCGTTTG 120
|||||
DB 206 TGGCGGAAATCTCGGTTTACGTTTATGCGCGGAAATACACGGGTTTAC-GTTTTGG 264
|||||
QY 121 CGGGAATATCACGGATTATGTTTTCGTTGAGAAATACGAGTTTACTTTTCTCAAT 180
|||||
DB 265 CGGGAATATCACGGTTTATGTTTTTAAAGAAATTACGTTTACTTTTCTCAAT 324
|||||
QY 181 TTCAATGCTTGATATTTAAAGAAATTTGAGAAATATTAATTTTAAATTTGGTTTGA 240
|||||
DB 325 TTCAATGCTTGATATTTAAAGAAATTTGAGAAATATTAATTTTAAATCGGCTTGA 384
|||||
QY 241 TGTGTTGTTAAACCTAAATTGGCATTTGGTTTAGAGATTTAGTTGGTTTATTCAATT 300
|||||
DB 385 TGTGTTGGCTAAACTTAAATTGGCATTTGGTTTAGAGATTTAGTTGGTTTATTCAATT 444
|||||
QY 301 TACAAATTTGATGGTTAAATTGGATAAACCATGAAACCATTAACCATTAACCTAAC 360
|||||
DB 445 TACAAACCTTGAAGGTTAAATTGGTTAAACCATTTGGAACCATTAACCATTAACCCACC 504
|||||
QY 361 TCATTTTACTCATCAACCAATTGACT 387
|||||
DB 505 TTATTTTACTCATCAACCAATTGACT 531
|||||

RESULT 8
CC055323
LOCUS
DEFINITION
  SALK_093737.53.25.x Arabidopsis thaliana TDNA insertion lines
  Arabidopsis thaliana genomic clone SALK_093737.53.25.x, genomic
  survey sequence.
ACCESSION
  CC055323
VERSION
  CC055323.1 GI:29474987
KEYWORDS
  GSS.
SOURCE
  Arabidopsis thaliana (thale cress)
ORGANISM
  Arabidopsis thaliana
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
  ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
  1 (bases 1 to 363)
  Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadzinab
  ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
  , Zimmerman,J. and Ecker,J.R.
  A Sequence-Indexed Library of Insertion Mutations in the
  Arabidopsis Genome
  Unpublished
  Contact: Joseph R. Ecker
  Salk Institute Genomic Analysis Laboratory (SIGAL)
  The Salk Institute for Biological Studies
  10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
  Tel: 858 453 4100 x1752
  Fax: 858 558 6379
  Email: ecker@salk.edu
  This is single pass sequence recovered from the left border of
  TDNA.
Class: TDNA tagged.
  Location/Qualifiers
    1..363
      /organism="Arabidopsis thaliana"
      /mol_type="genomic DNA"
      /strain="Columbia 0"
      /db_xref="taxon:3702"
      /clone="SALK_093737.53.25.x"
      /clone_lib="Arabidopsis thaliana TDNA insertion lines"
      /note="PCR was performed on Arabidopsis thaliana lines
      each of which contains one or more TDNA insertion
      elements. The resultant fragment for each line was
      directly sequenced to determine the genomic sequence at
      the site of insertion. Details of the protocols used can
      be found at http://signal.salk.edu/tdna_protocols.html"
BASE COUNT
  115 a 58 c 57 g 132 t

```

ORIGIN

Query Match	5.7%;	Score 292.4;	DB 29;	Length 363;
Best Local Similarity	92.2%;	Pred. No. 9.8e-27;		
Matches 319;	Conservative	0;	Mismatches 26;	Indels 11;
Gaps 1;				

QY 143 TTTTGGTGGAATAATACGAGTTTACTTTTCTCAATTTCAATGCTGTGTAATTTAAGA 202

Dd 1 TTTTGGTGGAATAATACGAGTTTACTTTTCTCAATTTCAATGCTGTGTAATTTAAGA 60

Qy 203 A-ATTTCGAAAAATATTAAATTTTATTAAATTCGTTTAGATGCTGGTTAAACCTAAATT 261

Db 61 ANATTTTCGAAAAATATTGATTTTCTTAAATTCGTTTAGATGCTGGTTAACTTAAATT 120

Qy	262	GCATTCGGTTTAGAGATTTT	TAGTTGGTTT	TATTCAA	TTTTACAAATTT	GATCGGTTTAA	321
Db	121	GATATTCGGTTTAGAGATTTT	TATTTGGTTTAA	TTCAA	TTTTACAAATTT	GTCGGTTTAA	180

Qy	381	TCGATAAACCATGGAAACCATTAAACCACTTAACTCATTTTACTCATCAACCAA
Db	240	TGGGTAAACCATTGAAACCACTTAACCTTACAACCCAACTCATTTTACTCTCAACCAA

Qy	382	TTTGACTCATCAACTCATTGTGACTCATCAACTCATTTTGAGTCAGAAAAATTTCAACTCAATTAG	441
Db	241	TTGACCCCAATAACTCATTGTGACCCATCAACTCATTTTGAGTCAGAAAAATTTCACTAAATTAG	300

Qy 442 GGTTCATGGGTTGAGTTGAGTTGAGTTGAGCCCATGAATTTGACCC 487
Db 301 GAATCATGGGTTGAGTTGAGTTGAGTTGAGCCCATGAATTCAGCC 346

RESULT 9
CNS00XC3

DEFINITION Arabidopsis thaliana genome survey sequence SP6 end of BAC T13K12 of TAMU library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.

VERSION AL094689.1 GI:5302844
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)

Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons; core eudicots; rosids; eurosids II: Brassicales; Brassicaceae; Arabidopsis.

AUTHORS
Salancubat, M., Choisine, N., Artiguenave, F., Brottier, P., Wincker, P.,
Samson, D., Saurin, W., Weissenbach, J. and Quétier, F.
JOURNAL
Unpublished

Genoscope.
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequençage :

```

FEATURES
  source
    1..441
    Location/Qualifiers
      - web : www.genoscope.cns.fr)

```

```

/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="EAXON:3702"
/mol_name="3702"

```

```
BASE COUNT      137 a 73 C 72 G 159 t
          137 a 73 C 72 G 159 t
/clone_lib="TAMU"
/notes="end : SP6"
```

Query Match	5.2%	Score 266;	DB 29;	Length 441;
Best Local Similarity	81.2%;			
Best Local Similarity	81.2%;			
Pred. No.	1.7e-23;			
Pred. No.	1.7e-23;			

QY 160 ACGAGTTTACTTTTCTCAATTTTCATTGCTTGTATATTAAAGAAATTTGGAAAATATTTA 219

Db	48	ACGAGTTTAAATCTTCTCAATTTCAATTCGTTATATATTTAAGAAATTCGGAATAATTC	107
Qy	220	ATTTTATTTAAATTTGGTTTATAGATGTCTGGTTAAACCTAAATTTGGCATTTGGTTTAGAGATT	279
Db	108	ATTGTATTTAAATTTGGTTTATAGATGTCTGGTTAACTTAAATTCGCGCATTTGGTTTAGAGATT	167
Qy	280	TTAGTTGGTTTTATTCAATTTTACAAATTTGATGGGTTAATTCGATTAACCATTCGAAAC	339
Db	168	TTAGTTGGTTTAAATTTATTTTACAAAACCTTGATGGGTTAATTCGGTAA	215
Qy	340	CATTAAACCATTAACAACCTTAACCTCATTTACTCATCAAAACCAATTTGACTCATCACTCAATT	399
Db	216	----AAACCATTAACAACCAACTCATTTATTTCATCAAAACCAATTTGACTCAATAACTCATTT	271
Qy	400	TGACTCATCAACTCATTTTGAGTCAAAAATTTCAACTCATTTAGGGTTTCATGGTTTGAGTTG	459
Db	272	TGACCCATCAACTCATTTTGAGTTAATATTTTCAACTCATTTAGGGTTTCATGGTTTGAGTTG	331
Qy	460	AGTTGAGTTTGACCCATGAATTTTGGACCCATTTTGGACCCCTACATATGATCATTAAGTTA	519
Db	332	AGTTGAGTTTGACCCATGAATTTTGATCCATTTTGACACCCCTAGGCCCGTCTCATGAGA	391
Qy	520	ATAATCAAAAAATTACATTATGATAAATTTTTTAAACGGAAATTTGTTTT	563
Db	392	TTTTAAAAACGACACAGTCATCGTTTAGCAACACCACTGAATATTTTT	435

RESULT 10
BH902224

LOCUS BH902224 319 bp DNA linear GSS 04-SEP-2002
 DEFINITION SALK_091482.54.20.x *Arabidopsis thaliana* TDNA insertion lines
 SALK_091482.54.20.x, genomic
 survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE AUTHORS	TITLE	JOURNAL COMMENT
----------------------	-------	--------------------

JOURNAL
COMMENT

COMMENT

FEATURES

sources

BASE COUNT

Db	313	TTTTGGCGGAAAAAATTACGGAATTTACGTTTGGCGGTAAATTCACGGGTTTACTTTTT	254
Qy	60	TTGGCGGAAAAATCTCGGGTTTACGTTTGGCGGAAAAATTCACGGGTTTACTTTTTTTT	119
Db	253	TTGGCGGAAAAATTACAGGTTTACGTTTGGAGGAAAAATTACGGGTTTACGTTTGTG	194
Qy	120	CGCGAAAAATCACGATTTA-TGTTTTTGGTGGAAAAATACGAGTTTACTTTTTCTCA	178
Db	193	GTGAAAAATCACGTTTACTTTTTTGGTGGAAAAATACGAGTTTACTTTCTTCA	134
Qy	179	ATTTCATTGCTGTATATTTAAGAAATTTGAAAAATTAATTAATTTATTAATGCTTCA	238
Db	133	ATTTCATCGCTGTATATTTAAGAAATTTAGGAAATTAATGATTTATTAATGCTTCA	74
Qy	239	GATGTGTTGGTTAAACCTTAAATTTGGCATTTGGTTTAGAGATTTAGTTGGTTTATCAAT	298
Db	73	GATGTGATGGTTAAACCTTAAATTTGGCATTTGGTTTAGAGATTTAGTTGGTTTATCAAT	14
Qy	299	TTTACAAAATTG 311	
Db	13	TTTACAAAATTG 1	
RESULT 12			
LOCUS	BX289638		
DEFINITION	Arabidopsis thaliana T-DNA flanking sequence GK-428H08-018077,		
ACCESSION	BX289638		
VERSION	BX289638.1	GI:28888634	
KEYWORDS	GSS.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
REFERENCE	1	Strizhov,N., Li,Y., Rosso,M., Viehoveer,P., Dekker,K., Saeidler,H. and Weisshaar,B.	
AUTHORS			
TITLE		A pipeline for automated high-throughput generation of FSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines	
JOURNAL		Unpublished	
REFERENCE	2	Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.	
AUTHORS			
TITLE		A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics	
JOURNAL		Unpublished	
REFERENCE	3	(bases 1 to 367)	
AUTHORS		Strizhov,N., Rosso,M., Li,Y. and Weisshaar,B.	
TITLE		Direct Submission	
JOURNAL		Submitted (07-MAR-2003) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany	
COMMENT		This sequence is recovered from the left border of the T-DNA. It indicates an insertion close to or within gene At3g28010. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/.	
FEATURES		Location/Qualifiers	
source		1..367	
		/organism="Arabidopsis thaliana"	
		/mol_type="genomic DNA"	
		/strain="Columbia 0"	
		/db_xref="taxon:3702"	
		/clone="GK-428H08-018077"	
		/clone_lib="Arabidopsis thaliana T-DNA insertion lines"	
		/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector PAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR	
BASE COUNT	222	a 99 c 51 g 156 t	
ORIGIN			
Query Match		5.1%; Score 259.8; DB 28; Length 319;	
Best Local Similarity		95.7%; Pred.No.1.1e-22;	
Matches	267;	Conservative 0; Mismatches 12; Indels 0; Gaps 0;	
Qy	229	AAATTGGTTAGATGTGGTTGTTAAACCTAAATTTGGCATTTGGTTTAGAGATTTAGTTGGT	288
Db	1	AAATTGGTTAGATGTGGTTGTTAAACCTAAATTTGGCATTTGGTTTAGAGATTTAGTTGGT	60
Qy	289	TTTATTCATTTTACAAAATTTGATGGTTAAATTTGGATTAACCATGGAACCATTAACCA	348
Db	61	TTTATTTAAATTTTACAAAATTTGATGGTTAAATTTGGTAAACCATGGAACCATTAACCA	120
Qy	349	TTACACACCTAACTCAATTTTACTCAACCAACCAATTTGACTCACTCACTCACTCACTC	408
Db	121	ATACACCACTCACTTTTACTCACTCAACCACTTTGACTCACTCACTCACTCACTC	180
Qy	409	AACTCAATTTGAGTCAAAATTTCACTCACTTTAGGGTTTCACTGGTTGAGTTGAGTT	468
Db	181	AACTCAATTTGAGTCAAAATTTCACTCACTTTAGGGTTTCACTGGTTGAGTTGAGTT	240
Qy	469	GACCCATGATTTTGCACCATTTTGCACCCCTTACATAT 507	
Db	241	GACTCATGAATTTTGACCCATTTTGACACCCCTTACATAT 279	
RESULT 11			
BH237177/c			
LOCUS	BH237177	528 bp DNA linear GSS 13-NOV-2001	
DEFINITION	AUIIA63TR AUII Arabidopsis thaliana genomic clone AUIIA63, genomic survey sequence.		
ACCESSION	BH237177		
VERSION	BH237177.1	GI:16907548	
KEYWORDS	GSS.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
REFERENCE	1	(bases 1 to 528)	
AUTHORS		Town,C.D., Whitelaw,C.A., Pai,G., Van Aken,S.E., Utterback,T.V., Feldblyum,T.V. and Fraser,C.M.	
TITLE		Survey sequencing of Arabidopsis thaliana BAC T2P24	
JOURNAL		Unpublished	
COMMENT		Other GSSs: AUIIA63TF Contact: Chris Town	
FEATURES		Location/Qualifiers	
source		1..528	
		/organism="Arabidopsis thaliana"	
		/mol_type="genomic DNA"	
		/strain="Columbia"	
		/db_xref="taxon:3702"	
		/clone="AUIIA63"	
		/clone_lib="AUII"	
		/note="Vector: pHO52; Site 1: BstXI; 2-3 kb sheared BAC DNA inserted into pHO52 using BstXI linkers"	
BASE COUNT	222	a 99 c 51 g 156 t	
ORIGIN			
Query Match		4.5%; Score 231.4; DB 28; Length 528;	
Best Local Similarity		87.9%; Pred.No.2.9e-19;	
Matches	275;	Conservative 0; Mismatches 36; Indels 2; Gaps 2;	
Qy	1	TTTTTGGCGGAAAAATCTCGGGTTTACGTTTGGCGGAAAAATCTCGGTTTACGTTT	59

were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequences were processed for submission. T-DNA derived sequences were removed"

```
BASE COUNT      113 a   72 c   55 g   127 t
ORIGIN

Query Match      4.2%; Score 216.6; DB 29; Length 367;
Best Local Similarity 90.6%; Pred. NO. 2.2e-17;
Matches 242; Conservative 0; Mismatches 24; Indels 1; Gaps 1;

QY 245 TTGGTTAAACCTAAATGGCATTGGTTTAGAGATTTTGGTGGTTT-ATTCAATTTTAC 303
Db 54 TTGGTTAAAGCTTAAATGGCATTGGGTTAAAGATATTTATTGGATTGCATTCATTTTAC 113

QY 304 AAAATTTGATGGGTTAAATGGATAAACCATGGAACCATTAACCATTAACAACCTAACTCA 363
Db 114 AAAATTTGATAGTTAAATGGGTTAAACCATTTGAACCATTAACCATTAACAACCTCA 173

QY 364 TTTTACTCATCAACCAATTTGACTCATCACTCATTTTGGTCTCATCACTCATTTTGAGTCA 423
Db 174 TTTTACTCATCAACCAATTTGACTCATCACTCATTTTGGTCTCATCACTCATTTTGAGTCA 233

QY 424 AAAATTTCACTCATTTAGGTTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTG 483
Db 234 AAAATTTTAACTCATTTAAAGTTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTG 293

QY 484 ACCCATTTTGACACCCCTCATATGAT 510
Db 294 ACCCATTTTGACACCCCTCATATATAT 320
```

```
RESULT 13
B77189/c
LOCUS      B77189 598 bp DNA linear GSS 16-JAN-1998
DEFINITION T3117R TAMU Arabidopsis thaliana genomic clone T3117, genomic
survey sequence.
ACCESSION B77189
VERSION B77189.1 GI:2773828
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 598)
AUTHORS Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
,J.C.
A BAC End Sequence Database for Identifying Minimal Overlaps in
Arabidopsis Genomic Sequencing. Update 3
JOURNAL Other GSSs: T3117TF
COMMENT Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 598.
FEATURES
source
1..598
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T3117"
/sex="hermaphrodite"
/clone_lib="TAMU"
```

```
/note="Vector: BclOBACII; Site_1: HindIII; Site_2: HindIII
; Produced by Rod Wing"
BASE COUNT      221 a   83 c   117 g   177 t
ORIGIN

Query Match      3.8%; Score 194.2; DB 28; Length 598;
Best Local Similarity 66.1%; Pred. No. 1.1e-14;
Matches 326; Conservative 0; Mismatches 163; Indels 4; Gaps 3;

QY 270 TTTAGAGATTTTGGTGGTTTATTCAATTTTACAAAATTTGATGGTTAAATGGATAAA 329
Db 598 TTTAAAGATTTTGGTGGTTTAAATTTCAATTTTACAAAATTTGATGGTTAAATGGTAA 539

QY 330 CCATGGAACCACTTAA-CCATTACACCTAACTCACTTTTACTCATCAACCACTTCACTC 388
Db 538 CCATTAACCACTTAAACCCATTACACCACTCACTTTTACTCATCAACCACTTCACTC 479

QY 389 ATCAACTCATTTGACTCATCACTCATTTTGAAGTCAAAAATTTCACTCATTTAGGGTTCA 448
Db 478 ATCAACTCATTTGACCCATCACTCATTTTGAAGTCAAAA-TCACCCATTTAGGGTTCA 421

QY 449 GGGTTGAGTTGAGTTGAGTTGAGTTGAGTTTGGCCCATTTTGGCCCATTTTGGCCCAT 508
Db 420 GGGTTGAGTTGAGTTGAGTTGAGTTGAGTTTGGCCCATTTTGGCCCATTTTGGCCCAT 361

QY 509 ATCATAGTTTAATTAATCAAAAATTTACTATTGATACTTTTAAAGGAATTTTGTGAAG 568
Db 360 GGGCTTACTTTA-ATTTTTTTTTTTTTTTTGTAAAATTTAACTTAAAGTTTGTAGAA 302

QY 569 TTTCAATTTACGTTATTTATATAAAAAACATAATGCAAAAGTACTAATGTATAGTTATTT 628
Db 301 AGTAACAAAATATTTCCACCGATTATGAATTTAATCCTATTAACTATCGTAAACATC 242

QY 629 TTATTTTAAATGAACAAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 688
Db 241 TTAACCTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 182

QY 689 CTATATTTCAATTAATGACCAACCACTCATATTTCTCTCTTCTTCTTCTTCTTCTTCT 748
Db 181 TTGGTTTCTTGGTAAACAATCAAGGCTCTCTCTGACATTTGGCCCACTTAATGATAT 122

QY 749 CTGCTCTCTGTAT 761
Db 121 GTGGCCTTCTGTCT 109

RESULT 14
AV830052
LOCUS      AV830052 514 bp mRNA linear EST 01-APR-2002
DEFINITION AV830052 RAFL9 Arabidopsis thaliana cDNA clone RAFL09-61-G03 5',
mRNA sequence.
ACCESSION AV830052
VERSION AV830052.1 GI:19872112
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 514)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,
Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.
and Shinozaki,K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL Unpublished
COMMENT Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: mseki@rtc.riken.go.jp
```

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLIC-1 vector (Carlini et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pluscript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES Location/Qualifiers
 source 1. .514
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAF109-61-G03"
 /dev_stage="plants at various developmental stages from germination to mature seeds"
 /lab_host="DH10B"
 /clone_lib="RAF19"
 /note="Site 1: BamHI; Site 2: SalI; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"
 BASE COUNT 163 a 88 c 123 g 137 t 3 others
 ORIGIN
 Query Match 3.6%; Score 185.8; DB 9; Length 514;
 Best Local Similarity 98.9%; Pred. No. 1.3e-13;
 Matches 187; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1983 TCAGATAAAGTTGTAAGATGACAAAGAAATGAGAGAGTGGCGAAGTACGAGTT 2042
 Db 77 TCTGATAAAGTTGTAAGATGACAAAGAAATGAGAGAGTGGCGAAGTACGAGTT 136
 QY 2043 GGTGCGACAAATAGTCAAGAACCTTGTCTAGGTTAAGTTGCGAGGACACAGACCT 2102
 Db 137 GGTGCGACAAATAGTCAAGAACCTTGTCTAGGTTAAGTTGCGAGGACACAGACCT 196
 QY 2103 GGTGATAATGAGCCATCAAAATTTAGTCTAAGAGTACATACTTAAGAACACAGATGTT 2162
 Db 197 GGTGATAATGAGCCATCAAAATTTAGTCTAAGAGTACATACTTAAGAACACAGATGTT 256
 QY 2163 GATCAGGTA 2171
 Db 257 GATCAGATA 265

RESULT 15
 BX291485/c
 LOCUS
 DEFINITION 454 bp DNA linear GSS 07-MAR-2003
 Arabidopsis thaliana T-DNA flanking sequence GK-463G11-018953,
 genomic survey sequence.
 ACCESSION BX291485
 VERSION BX291485.1 GI:28890481
 KEYWORDS GSS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
 AUTHORS and Weissshaar,B.
 TITLE A pipeline for automated high-throughput generation of FSTs
 (flanking sequence tags) from Arabidopsis thaliana T-DNA
 transformed lines
 JOURNAL Unpublished
 REFERENCE 2
 AUTHORS Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weissshaar,B.
 TITLE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
 for flanking sequence tag based reverse genetics
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 454)
 AUTHORS Li,Y., Strizhov,N., Rosso,M. and Weissshaar,B.
 TITLE Direct Submission
 JOURNAL Submitted (07-MAR-2003) Weissshaar B., Max-Planck-Institut fuer

COMMENT

Zuechtungsforschung, Carl-von-Linne-Meg 10, Koeln, 50829, Germany
 This sequence is recovered from the left border of the T-DNA. It
 indicates an insertion within the locus defined by clone P9C16. The
 sequences are generated at the MPI for Plant Breeding Research in
 the context of the GABI-Kat project. GABI-Kat is part of the German
 Plant Genomics program designated 'GABI'. Information on line
 availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES

source

Location/Qualifiers
 1. .454

/organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="GK-463G11-018953"
 /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
 /note="PCR was performed on DNA from Arabidopsis thaliana
 plants (T1) which were transformed with the T-DNA from
 vector PAC161. The lines contain one or more T-DNA
 insertions. The DNA fragment(s) resulting from the PCR
 were directly sequenced to determine the genomic sequence
 flanking the insertion. Sequences displaying significant
 similarity to the A. thaliana nuclear genome sequence were
 processed for submission. T-DNA derived sequences were
 removed"

BASE COUNT 171 a 91 c 45 g 147 t

ORIGIN

Query Match 3.4%; Score 173.6; DB 29; Length 454;
 Best Local Similarity 77.2%; Pred. No. 4.2e-12;
 Matches 227; Conservative 0; Mismatches 59; Indels 8; Gaps 1;
 QY 179 ATTTCATTGCTGTATATTTAAGAAATTTGGAATAATTAATTTTAAATTTGTTTA 238
 Db 306 AATAAAATTGATGGTTGGTGAAGCCATTGTAAAAGATTGATTATTTCAATTGATT 247
 QY 239 GATGTGTTGGTTAAACCTAAATTTGCATTGGTTTGAAGATTTTAGTTGGTTTATTCAT 298
 Db 246 GAGGTTTTGGTTAAACTCAATTTGATATGATTTAGATATTTAAGTTGGATTAATTCAT 187
 QY 299 TTTACAAAATTTGATGGTTTAAATTTGATAAAACCATGGAAACCATTAACCATTAACACCTA 358
 Db 186 TTAGCCAAAATTTGGTGGTTAAATGGTAAACCAT-----TAAACCATTTGAACCCA 135
 QY 359 ACTCATTTTACTATCAACAAACCAATTTGACTATCACTCATTTGACTCATCACTCATTTG 418
 Db 134 ACTCATTTTATCCATCTAACCATTTTGACCAATAACTCATTTGAACATATCAATTCATTG 75
 QY 419 AGTCAAAAATTTCAACTCATTTAGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGACC 472
 Db 74 AGTGA AAAATTTTCATCTCTATTAGGGTTTCATGGTTGAGTTGAGTTGAGTTGATC 21

Search completed: October 21, 2003, 04:26:39

Job time : 9747 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2003, 04:26:42 ; Search time 1236 Seconds

(without alignments)
11161.152 Million cell updates/sec

Title: US-09-824-735-1

Perfect score: 5144

Sequence: 1 ttttggggggaatactcg.....acgttttgatcccaacttaa 5144

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584790

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/BCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5144	100.0	5144	9	US-09-824-735-1
2	414.6	8.1	2000	10	US-09-938-842A-2784
3	401.4	7.8	2000	10	US-09-938-842A-5327
4	371.2	7.2	2000	10	US-09-938-842A-4657
5	152.4	3.0	2000	10	US-09-938-842A-5099
6	97.6	1.9	4004	10	US-09-878-574-3933
7	95.6	1.9	1468	12	US-10-292-408-8
8	91.6	1.8	3673778	12	US-10-312-841-1
9	88	1.7	742	9	US-09-770-149-51
10	83.4	1.6	3673778	12	US-10-312-841-2
11	82.2	1.6	1326	10	US-09-938-842A-72
12	81.8	1.6	15548	12	US-10-311-455-2128
13	80.4	1.6	15367	12	US-10-311-455-157
14	77.4	1.5	3673778	12	US-10-312-841-2
15	76.6	1.5	7306	12	US-10-311-455-1610
16	75.2	1.5	1326	10	US-09-938-842A-2313

17	74.6	1.5	1470	10	US-09-938-842A-1978
18	74.2	1.4	5893	12	US-10-311-455-832
19	73.8	1.4	464	9	US-09-770-444-356
20	73.8	1.4	1290	10	US-09-938-842A-1831
21	73.8	1.4	6314	12	US-10-240-452-14
22	72.4	1.4	1266	10	US-09-938-842A-1288
23	72.4	1.4	18218	12	US-10-311-455-1922
24	72.2	1.4	6030	12	US-10-240-453-186
25	72.2	1.4	6030	14	US-10-239-676-164
c 26	71.4	1.4	461	12	US-09-814-353-17724
c 27	71.4	1.4	2000	10	US-09-938-842A-3403
28	70.8	1.4	7002	12	US-10-311-455-783
29	70.8	1.4	10480	12	US-10-311-455-2173
30	70.4	1.4	16724	12	US-10-311-455-1063
31	70.4	1.4	16724	12	US-10-240-485-89
32	70.2	1.4	393	10	US-09-960-352-4582
33	70	1.4	6419	12	US-10-311-455-240
34	70	1.4	17183	12	US-10-311-455-460
35	69.2	1.3	5689	12	US-10-240-453-100
36	69.2	1.3	5689	14	US-10-239-676-90
c 37	69.2	1.3	6590	12	US-10-311-455-1449
38	69.2	1.3	7143	12	US-10-311-455-956
39	69	1.3	480	10	US-09-960-352-5301
40	68.6	1.3	1281	10	US-09-938-842A-70
41	68.2	1.3	6341	12	US-10-311-455-1617
42	68.2	1.3	9707	12	US-10-311-455-1393
43	68	1.3	9539	12	US-10-240-453-53
44	68	1.3	9539	14	US-10-239-676-51
45	68	1.3	12705	12	US-10-311-455-122

ALIGNMENTS

RESULT 1

US-09-824-735-1
; Sequence 1, Application US/09824735
; Patent No. US20020095032A1
; GENERAL INFORMATION:
; APPLICANT: ZHU, JIAN-KANG
; APPLICANT: LIU, JIPING
; APPLICANT: ISHITANI, MANABU
; APPLICANT: HALFTER, URSULA
; APPLICANT: KIM, CHEOL-SOO
; TITLE OF INVENTION: PROTEINS AND DNA RELATED TO SALT TOLERANCE IN PLANTS
; FILE REFERENCE: 205645US20
; CURRENT APPLICATION NUMBER: US/09/824,735
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 60/824,735
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 5144
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2004)..(2168)
; OTHER INFORMATION:
; NAME/KEY: CDS
; LOCATION: (2255)..(2317)
; OTHER INFORMATION:
; NAME/KEY: CDS
; LOCATION: (2416)..(2487)
; OTHER INFORMATION:
; NAME/KEY: CDS
; LOCATION: (2573)..(2680)
; OTHER INFORMATION:
; NAME/KEY: CDS
; LOCATION: (2778)..(2852)
; OTHER INFORMATION:
; NAME/KEY: CDS

Sequence 1978, Ap
Sequence 832, App
Sequence 356, App
Sequence 1831, Ap
Sequence 14, Appl
Sequence 1288, Ap
Sequence 1922, Ap
Sequence 186, App
Sequence 164, App
Sequence 17724, A
Sequence 3403, Ap
Sequence 783, App
Sequence 2173, Ap
Sequence 1663, Ap
Sequence 89, Appl
Sequence 4582, Ap
Sequence 240, App
Sequence 460, App
Sequence 100, App
Sequence 90, Appl
Sequence 1449, Ap
Sequence 956, App
Sequence 5301, Ap
Sequence 70, Appl
Sequence 1617, Ap
Sequence 1393, Ap
Sequence 53, Appl
Sequence 51, Appl
Sequence 122, App

[illegible]

Db	2821	CGGATTTCCGACTCAGTGCATTCGCTTCAGGAAGTAAGTGTCTTTATCTCTGCTTCAGCAG	2889
Qy	2881	TCCTCATTACGTGGTCATTAACTTCGTATATACCTCAATCAGGAGTAGAACTTCCTGCGTAC	2940
Db	2881	TCCTCATTACGTGGTCATTAACTTCGTATATACCTCAATCAGGAGTAGAACTTCCTGCGTAC	2940
Qy	2941	CACATGTGGAACTCCGAACATATGATGCTCCAGAGGTACTTAGTGGACAGGGTTACGATGG	3000
Db	2941	CACATGTGGAACTCCGAACATATGATGCTCCAGAGGTACTTAGTGGACAGGGTTACGATGG	3000
Qy	3001	TTCCAGCAGCTGATATTTTGGTCTTCGCGGGTTATTCCTTTTCGTATATATTTGGCTGCGATATTT	3060
Db	3001	TTCCAGCAGCTGATATTTTGGTCTTCGCGGGTTATTCCTTTTCGTATATATTTGGCTGCGATATTT	3060
Qy	3061	ACCTTTTCCGAGACGGATCTTCAGAGGTGTACAGAAAAGTAAGTAACATATCTTTTCGG	3120
Db	3061	ACCTTTTCCGAGACGGATCTTCAGAGGTGTACAGAAAAGTAAGTAACATATCTTTTCGG	3120
Qy	3121	GAAGAAATCATGAATTCCTTGTCATGGCTTTTGTCAAACCGTTTATTTGATTTGGTTTTTGC	3180
Db	3121	GAAGAAATCATGAATTCCTTGTCATGGCTTTTGTCAAACCGTTTATTTGATTTGGTTTTTGC	3180
Qy	3181	AAATTCACCCCTTAGTTTGGAGCTTTTACATTTGTTATTTACAGATAAATGCAGCAGAGT	3240
Db	3181	AAATTCACCCCTTAGTTTGGAGCTTTTACATTTGTTATTTACAGATAAATGCAGCAGAGT	3240
Qy	3241	TTTTCTGTCCACCGTGTGTTTTCCGCAAGTGAAGTTTTTAAATACATAGGATACTTTCACC	3300
Db	3241	TTTTCTGTCCACCGTGTGTTTTCCGCAAGTGAAGTTTTTAAATACATAGGATACTTTCACC	3300
Qy	3301	CCAAATCCCAAAACAGTGAATTTTGTGTTCTCTCCTAGCTATCAGGTTTGTGGTAT	3360
Db	3301	CCAAATCCCAAAACAGTGAATTTTGTGTTCTCTCCTAGCTATCAGGTTTGTGGTAT	3360
Qy	3361	ATTTAAATGTTCTAGTAATATATCTGTTTATCTATATATGTTTCTCAATTAGACACGGT	3420
Db	3361	ATTTAAATGTTCTAGTAATATATCTGTTTATCTATATATGTTTCTCAATTAGACACGGT	3420
Qy	3421	ATTTCAAAATTCAGGAATTCAGAAAGATCCCTGGTTAGATTAAATATGTCCCTATACGA	3480
Db	3421	ATTTCAAAATTCAGGAATTCAGAAAGATCCCTGGTTAGATTAAATATGTCCCTATACGA	3480
Qy	3481	GCAAGGGAAGAAGAAGTGAATTCGATGATATTCGTGCAGTTTTTGTAGTAATTTAG	3540
Db	3481	GCAAGGGAAGAAGAAGTGAATTCGATGATATTCGTGCAGTTTTTGTAGTAATTTAG	3540
Qy	3541	GTTTGTGCTTCGCCCTTCATTATATGCTCTTTGCTGGTCAATTCGATTTAAATGTTAAGAT	3600
Db	3541	GTTTGTGCTTCGCCCTTCATTATATGCTCTTTGCTGGTCAATTCGATTTAAATGTTAAGAT	3600
Qy	3601	CTCTTTAGGAAGTTTGGATGACCAAGAGAGTGTGTTGCTACAGGATAGAACAAAATAGT	3660
Db	3601	CTCTTTAGGAAGTTTGGATGACCAAGAGAGTGTGTTGCTACAGGATAGAACAAAATAGT	3660
Qy	3661	AGGCATGTGTTAGTTACCAAAACCTGTAAACTGCTTCTTATTCAAATTCGCCAAACCATAG	3720
Db	3661	AGGCATGTGTTAGTTACCAAAACCTGTAAACTGCTTCTTATTCAAATTCGCCAAACCATAG	3720
Qy	3721	ACCTTAGGAAGACTTAGATTACAGAGAAATCTCTATTCGACCAAAAACCCCTAGACA	3780
Db	3721	ACCTTAGGAAGACTTAGATTACAGAGAAATCTCTATTCGACCAAAAACCCCTAGACA	3780
Qy	3781	AAATCCAGATACCCCTAGGCTAAATTCAGATGTTCTATGTCACCTATCAATATATATCTCC	3840
Db	3781	AAATCCAGATACCCCTAGGCTAAATTCAGATGTTCTATGTCACCTATCAATATATATCTCC	3840
Qy	3841	TGCTAAAGATGTCTTAGCTTTGGTATAGCCTAGATATATAGATACCGTAATTTCTAAT	3900
Db	3841	TGCTAAAGATGTCTTAGCTTTGGTATAGCCTAGATATATAGATACCGTAATTTCTAAT	3900
Qy	3901	GCATATTTAGAAATGTTTTACTTAAACTCAGSCTCTTGCTCTTCTTAAACCTTGACTTCAA	3960
Db	3901	GCATATTTAGAAATGTTTTACTTAAACTCAGSCTCTTGCTCTTCTTAAACCTTGACTTCAA	3960

	Query Match	1.9%	Score 97.6;	DB 10;	Length 404;
	Best Local Similarity	64.7%;	pred. No. 2.6e-07;		
	Matches 145;	Conservative	0;	Mismatches 79;	Indels 0;
	Gaps	0;			
Qy	2916	ATCAGGAGTAGAACCTTCGGGTACACATGTGGAACTCCGAACTATGTAGCTCCAGAGG	29175		
Db	78	AAcAGGAGGATGAACCTTCcGGcACTcCTTGTGTGGAACTCCAAATATATGTTGTCTCTGAGG	137		
Qy	2976	TACTTAGTCGACAGGGTTACGATGGTTTCACGACCTGATATTGGTCTTCGGGGGTATTTC	3035		
Db	138	TGCTTAATGATAGAGGGTTATGTGGTTTCTACATCTGATATCTGCTCTGGAGTCAATTC	197		

```

US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication NO. US20030186277A1
GENERAL INFORMATION:
APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von b
; FILE REFERENCE: E01/1208/WO
CURRENT APPLICATION NUMBER: US/10/
CURRENT FILING DATE: 2002-12-30
NUMBER OF SEQ ID NOS: 2
SEQ ID NO 1
LENGTH: 3673778
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically tre
FEATURE:
NAME/KEY: unsure
LOCATION: (3294164)

```

US-10-312-841-1

Query Match 1.8%; Score 91.6; DB 12; Length 3673778;
Best Local Similarity 43.2%; Pred. NO. 0.00012;
Matches 546; Conservative 0; Mismatches 709; Indels 9; G

Qy	181	TTCAATTCGCTGTATAATTAAAGAAATTTGGAAAAAATTAATATTTTATAATTAATCGGTTAGA	240
Db	1713092	TTAATAATAATATAATTTATATATAAAAATAATAATATATATACGCTATTAATTTATATATTTTG	1713151
Qy	241	TGTGTGGTTAAACCTAAAATTGGCATTGGTTTAGAGATTTTAGTGTGGTTTATATTCAAATTT	300
Db	1713152	TGTGTGTGATATAATATATTTTATATATAATTTTATATATATAATTTTATATATGATATTT	1713211
Qy	301	TACAAAATTTGATGGGTTAAATGGATAAACCATGGAAACCATTAACCATTACAAACCTAAC	360
Db	1713212	TATATATATTTTATATATATAATTTTATATATATTTTATATATATTTTATATATTTATTTTTAT	1713271
Qy	361	TCATTTTACTCAUCAAAACCAATTTGACTCATCAACTCAATTTGACTCATCACTCAATTTGAG	420
Db	1713272	ATAATTTTATATATTTTATATATATGTTTATATATTTTATATATATATTTTATATATA	1713331
Qy	421	TCAAAATTTCAAATCATATAGGTTCAATGGGTTGAAGTTGAGTTGAGTTGACCAATGAAT	480
Db	1713332	TTTTATATATATTTTATATATATAATTTTATATATATTTTATATATATTTTATATATATAT	1713391
Qy	481	TTGACCAATTTTGACACCCCACATATGATCAATAAGTTAATAATCAAAAATTTACTATTTGA	540
Db	1713392	TTTTATATATATTTTATATATATTTTATATATATTTTATATATTTTATATATATATTT	1713451
Qy	541	TAACTTTTAAACCGAAATGTTTGTGAAGTTTCATTTACGTTATTTATATATATATATATATAT	600
Db	1713452	TTATATATATTTTATATATATAATTTTATATATATAATTTTATATATATTTTATATATATAT	1713511
Qy	601	ATGCCAAAGTACTAATGTATAGTTATTTTATTTTAAATAAGTCAAAAATTTACTGTAA	660
Db	1713512	TTTTATATATATAATTTATATATATTTTATATATAATTTTATATATATATTTTATATATA	1713571
Qy	661	TACTTTTTCATCTTATACAATTTTTTCTTATATTCAAITTAGCCCAACACACTCAAT	720
Db	1713572	TTTTATATATTTTATATATAATTTTATATATAATTTTATATATAATTTTATATATATTTATATA	1713631
Qy	721	TCATATTCCTTCCACATTCATTAATTTTCTGCTCTTGATATCTCTTATCATCTTCAAT	780
Db	1713632	TTTTTATATATTTTATATATAATTTTATATATAATTTTATATATTTTATATATATTTTATA	1713691
Qy	781	TGCCAAATTTTCANTGACATTTGCATCGTTACTTTGTGATGATGATTTTGTGAACATTTAA	840
Db	1713692	TATATTTTATATATTTATATATATAATAATATATATATATATATATATATATATATATATTA	1713751
Qy	841	TGTGARCTCCCAACCGACTATAGGACTATCAAGAAAGTTTTCAAACCTTTTAAAAAAGACT	900
Db	1713752	TATATATTTTATATATATTTTATATAATTTTATATATATTTTATATATTTTATATATATTT	1713811
Qy	901	TGAACCTTTAAAGCAATCCCAACCGACTATAGAACCATAGTCCCTCTCTGTGTGATGA	960
Db	1713812	TATATATTTTATATATATTTTAT	1713871
Qy	961	AGCTCTTCTCGTCGGCGGAAAACTTAGGCCATAAAAAGCCTCTTCCAACATCACCTAGTAT	1020
Db	1713872	ATAATTTTATATATATATATATAATTTTATATATATATATATATATTTTATATATATTTTAT	1713931
Qy	1021	ATTGACCGTGACCATCTTTTTTGACCATTTGCTTGTGGAATGAAACCGTCGATAAACCGTGT	- 1079
Db	1713932	ATATATATTTTATATATATTTTATATATATATTTTATATATATTTTATATATATATTTTATA	1713991
Qy	1080	--TATCACTTAGGCCAAATTTTTTCCCTAGTGTGTTTCCAAAAGCTCTCACGAAPCCCTTAT	1137
Db	1713992	TATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATTTTAT	1714051
Qy	1138	CGAACTTTTTTATATATACATTTGTAGCAT-----TGGAAAGTATCTTTGTATGCTTTG	1191
Db	1714052	ATATATTTTATATATATTTTATATATATATTTTATATATATTTTATATATATTTTATATATTTTATA	1714111

QY 1450 AAAATGAAGATTCAAAATGCAATGCTAGTCTTTTAAAGTTTAAT 1494
 Db 12715 TATATTTATATATATATATATATATATATATATATATATAT 12759

RESULT 13
 US-10-311-455-157
 ; Sequence 157, Application US/10311455
 ; Publication No. US20030143606A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OLEK, Alexander
 ; APPLICANT: FIEPENSROCK, Christian
 ; APPLICANT: BERLIN, Kurt
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
 ; FILE REFERENCE: 5013.1014
 ; CURRENT APPLICATION NUMBER: US/10/311.455
 ; CURRENT FILING DATE: 2002-12-16
 ; PRIOR APPLICATION NUMBER: PCI/EP01/07537
 ; PRIOR FILING DATE: 2001-07-02
 ; PRIOR APPLICATION NUMBER: DE 10032529.7
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: DE 10043826.1
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 2424
 ; SEQ ID NO 157
 ; LENGTH: 15387
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 US-10-311-455-157

Query Match 1.6%; Score 80.4; DB 12; Length 15387;
 Best Local Similarity 44.1%; Pred. No. 0.001;
 Matches 572; Conservative 0; Mismatches 706; Indels 18; Gaps 5;

QY 136 TTTATGTTTTTGGTGGAAAATACGAGTT---TACTTTTTCACATTCATTCGCTGT 192
 Db 9779 TTTGTTTTTGTAGTTTAAAGAAATGAAATTTGGGGTGTATTTTGTATTTTTTTTTTT 9838

QY 193 ATATTAAAGAAATTTGGAAAATATTAATTTTATTAATTTAGATGTTGGTTAA 252
 Db 9839 ACGTTTTTATAGTTTATAGATATTAAGTTTGGAAATTTATTTTATTTTGTGTTA 9898

QY 253 ACTAAATGGCATTCGTTAGAGATTTT-AGTTGGTTTATTCAAATTTACAAAATTTG 311
 Db 9899 TTTTATTGTTATTTTATTTATAGATATTAAGTTTGTAAATTTTATTTTATTTTGT 9958

QY 312 ATGGGTTAATTCGATAAAACCATGGAAACCAATTAACCACTAACCTCACTTTTACTC 371
 Db 9959 TTGTTTTTATCTATTTTATTTTATTTAGATATTAAGTTTGTAAATTTTATTTTATTT 10018

QY 372 ATCAACCAATTCATCACTCACTCATTTGACTCATCACTCATTTGAGTCAAAATTC 431
 Db 10019 TGTGTTTTTATTTATTTTATTTTATTTAGATATTAAGTTTGTAAATTTTATTTTAT 10078

QY 432 AACTCAATTAGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 491
 Db 10079 TTTTGTGTTATTTTATTTGTTATTTTATTTATTTATTTATTTATTTATTTATTTT 10138

QY 492 TGACACCCCTACATATGATCATAGTTAAATCAAAAATTTACTATTGATCACTTTTAA 551
 Db 10139 TATTTTGTGTTATTTTATTTGTTATTTTATTTATTTATTTATTTATTTATTTATTT 10198

QY 552 CGGAATGTTTGTAGTTTTCATTTACGTTATTTATTAATAAACAATATGCAAAAGTA 611
 Db 10199 TTTTATTTTGTGTTTATTTATTTGTTATTTTATTTATTTATTTATTTATTTATTT 10258

QY 612 CTAATGTATAGTTATTTTATTTTATTTTAAATGCAAAAATTTACTGTATATCTTTTCA 671
 Db 10259 ATTTTATTTTGTGTTATTTATTTGTTATTTTATTTATTTATTTATTTATTTATTT 10318

QY 672 TCTTATCAATATTTTCTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 731
 Db 10319 TTTAAATTTTATTTTGTGTTATTTTATTTATTTTATTTATTTATTTATTTATTTATTT 10377

QY 732 TTCACATTCATTTATTTCTGCTCTGTTATTCCTTCTTATCATCTTCAATTCGCAATTTT 791
 Db 10378 ---AAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 10434

QY 792 CATTGACATTCATCTGTTTACCTTTGATGATGATGATGATGATGATGATGATGATGAT 851
 Db 10435 TTTTGTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 10488

QY 852 ACCGACTATAGACTATCAAGAGTTTTCACAACTTTTAAAGAAAGATCTGCAACCTTAA 911
 Db 10489 --GATATTAAGTTTGTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTT 10546

QY 912 AGCAATCCCAACCGACTATAGAACCATAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 971
 Db 10547 TATTAGATATTAAGTTTGTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 10606

QY 972 TGCCGCGAAAATCTAGGCCATPAAAGCCTCTCAACATCACTACCTAGTATATTTGACCGTGA 1031
 Db 10607 TTTTATTTAGATATTAAGTTTGTAAATTTTATTTTATTTTATTTTATTTTATTTTATTT 10666

QY 1032 CCATCTTTTGTGACCATTTGCTTTGTGTAATGAACCGTGAACCGTGTATCACTTACGC 1091
 Db 10667 TTTTATTTATAGATATTAAGTTTGTAAATTTTATTTTATTTTATTTTATTTTATTTTAT 10726

QY 1092 CAAATTTTCCCTAGTGT---GTTTCCAAACTCTCAGAAATCTTATCGAACTTTTATA 1149
 Db 10727 TTAATTTTATTTATTTAGATATTAAGTTTGTAAATTTTATTTTATTTTATTTTATTT 10786

QY 1150 TATATCACTTTGTAGCAATGGAAAGTAICTTTGTGATGCTTTGTCTTAAACCTAGACATCC 1209
 Db 10787 TTGTTATTTTATTTATTTAGATATTAAGTTTGTAAATTTTATTTTATTTTATTTTATTT 10846

QY 1210 TTGTTCTCTGTTGTTTTTTTGAACCTTGTCTTGACTTAAATGAAAGTTTAAAAATTTGTAGTTA 1269
 Db 10847 TTAATGTTATTTTATTTATTTAGATATTAAGTTTGTAAATTTTATTTTATTTTATTTTAT 10906

QY 1270 AAAATAGAAAAATTTACTAAATTTGAGTTGCGATTAATCATAGTCTAGATAATTTGAAAAA 1329
 Db 10907 TTTTATTTGTTATTTTATTTATTTAGATATTAAGTTTGTAAATTTTATTTTATTTTATTT 10966

QY 1330 ATTAATAATAATTTTGAATACTATATGTTTATTTTAAAAATAATTTTACTATAATTTGA 1389
 Db 10967 TTAATTTTATTTGTTATTTTATTTATTTAGATATTAAGTTTGTAAATTTTATTTTATTT 11026

QY 1390 TAAAGTATATTAGATTTAGTTTTTTTTTCTTTTTTA 1425
 Db 11027 TGTGTTTTTATTTGTTATTTTATTTTATTTAGGTTTA 11062

RESULT 14

US-10-312-841-2/c
 ; Sequence 2, Application US/10312841
 ; Publication No. US20030186277A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Epigenomics AG
 ; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
 ; FILE REFERENCE: E01/1208/WO
 ; CURRENT APPLICATION NUMBER: US/10/312,841
 ; CURRENT FILING DATE: 2002-12-30
 ; NUMBER OF SEQ ID NOS: 2
 ; SEQ ID NO 2
 ; LENGTH: 3673778
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 ; NAME/KEY: unsure
 ; LOCATION: (379615)


```

; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-9550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13

Query Match 1.3%; Score 66.6; DB 2; Length 19124;
Best Local Similarity 50.8%; Pred. No. 0.00014;
Matches 159; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

Qy 1243 TTAAATCGAAGTTAAAAAATTGTAGTTTAAAAAATGAGAAAATTTTACTPAATTTTCAGTTTCGATT 1302
Db 15675 TTAATAAAAAAAAAAATATATATCATATAAACACAAAAAAGAAAAAATATATATAAT 15734
Qy 1303 AATCATAGCTAGATAAATTTGAAAAAATAAATAAATTTTGAAAAATCTATATGTTTA 1362
Db 15735 AAAAAATATATCATAAAAATAAAAAAATTTAAAAAATGTTTAAAAAATAATATATACA 15794
Qy 1363 TTTTTAAAAAATAATTACTAAATGATAAGTGATATTAGATTAGTTTTTTTTTTCTTTT 1422
Db 15795 TAAATFAAAAAAATTTATTTAAATAAAAAAATAAATAAATAAATAAATAAATAAT 15854
Qy 1423 TTAATTTTGAAACCTCACATTTATAATTGAAAGATTCAAAATGCAATGTTAGTTTT 1482
Db 15855 AAAAAAATAATAAATAAAAAAATTTAATTAATAAAAAAATAAATAAATAAATAAATAA 15914
Qy 1483 TAAAGTTTAAATCACCCAACGGAATAATTGCCCGAACGACTAATTCAGTCGTATACGG 1542
Db 15915 AATAAAAAAATAAAAAAATTTAATTAATAAAAAAATAAATAAATAAATAAATAAT 15974
Qy 1543 GTACAGTCAATA 1555
Db 15975 GCACATATACATA 15987

```

RESULT 3
US-08-998-416-595
; Sequence 595, Application US/08998416

Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philippssen, Peter

APPLICANT: Pohlmann, Rainer

APPLICANT: Steiner, Sabine

APPLICANT: Mohr, Christine

APPLICANT: Wendland, Jurgен

APPLICANT: Knechtle, Philipp

APPLICANT: Rebschung, Corinne

TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 1152

CORRESPONDENCE ADDRESS:

ADDRESS: No. 6239264artis Corporation

STREET: 3054 Cornwallis Road

CITY: Research Triangle Park

STATE: No. 6239264th Carolina

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/998,416

FILING DATE: 24-DEC-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 0016/97

FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 595:

SEQUENCE CHARACTERISTICS:

LENGTH: 658 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORGANISM: PAG1408RP

US-08-998-416-595

[illegible]

RESULT 4

US-09-056-075-1/c
 ; Sequence 1, Application US/09056075
 ; Patent No. 595368
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Eric A.
 ; APPLICANT: Bradshaw, Marite
 ; APPLICANT: Rood, Julian
 ; TITLE OF INVENTION: Expression System for Clostridium
 ; TITLE OF INVENTION: Species
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Quarles & Brady
 ; STREET: 1 South Pinckney Street
 ; CITY: Madison
 ; STATE: WI
 ; COUNTRY: US
 ; ZIP: 53701-2113
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/056,075
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Seay, Nicholas J.
 ; REGISTRATION NUMBER: 27386
 ; REFERENCE/DOCKET NUMBER: 960296.95238
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 608-251-5000
 ; TELEFAX: 608-251-9166
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 6243 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: misc.feature
 ; LOCATION: 3770..4013
 ; OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orit) from
 ; OTHER INFORMATION: plasmid RP4"
 ; US-09-056-075-1

Query Match 1.2%; Score 63.8; DB 2; Length 6243;
 Best Local Similarity 52.9%; Pred. No. 0.00042;
 Matches 137; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
 QY 523 ATCAAAATTAAGTATGATCACTTTTAAACGGAATGTTTGAAGTTTCATTACGTTA 582
 Db 1440 ATAAAAAATTTTTTTTCAACCTTTAAATAAAAAATATTTTATATTTTTTTTAA 1381
 QY 583 TTTATATAAAAAAACAATAAGCAAGTACTAATGATAGTATTTTATTTTATTTAATAAT 642
 Db 1380 TTTTATATTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1321
 QY 643 GCAAAATATTAAGTAACTATTTTCAATCTTTATCAATATTTTCTTATATTCATTA 702
 Db 1320 TATATTTTTTATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 1261
 QY 703 GCACCAACACCTACATTTTCATATTTCTCTTCACATCATATTTCTTGCTCTGAT 762
 Db 1260 TTTTATTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1201
 QY 763 CCTTCTTATCATCTTCATT 781
 Db 1200 TTTTATTTTACCCTCATT 1182

RESULT 5

US-08-487-826B-13/c
 ; Sequence 13, Application US/08487826B
 ; Patent No. 5993827
 ; GENERAL INFORMATION:
 ; APPLICANT: Sim, Kim L.
 ; APPLICANT: Chitnis, Chetan
 ; APPLICANT: Miller, Louis H.
 ; APPLICANT: Peterson, David S.
 ; APPLICANT: Su, Xin-zhaun
 ; APPLICANT: Wellem, Thomas E.
 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 ; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 ; NUMBER OF SEQUENCES: 45
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe Martens Olson & Bear
 ; STREET: 620 Newport Center Drive 16th Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/487,826B
 ; FILING DATE: 10-SEP-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Israel, Ned
 ; REGISTRATION NUMBER: 29,655
 ; REFERENCE/DOCKET NUMBER: NIH121.001CP1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 235-8550
 ; TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 19124 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; US-08-487-826B-13

Query Match 1.2%; Score 63.4; DB 2; Length 19124;
 Best Local Similarity 51.7%; Pred. No. 0.00065;
 Matches 168; Conservative 0; Mismatches 156; Indels 1; Gaps 1;
 QY 521 TAATCAAAATTAAGTATGATCACTTTTAAACGGAATGTTTGAAGTTTCATTACGT 580
 Db 15915 TCATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15856
 QY 581 TATTTATATATAAAAAACAATAAGTAACTAAATGTATAGTATTTTATTTTATTAATA 640
 Db 15855 TATTTAATATAATTTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15796
 QY 641 ATGCAAAATATTAAGTAACTTTTTCATTTTATCAATATTTTCTTATATTCAT 700
 Db 15795 ATGTATATATTTTTTTTAAACATTTTTTAAATTTTTTATTTTATGATATATTT 15736
 QY 701 TAGCCACCAACACCTACATTTTCATATTTCTCTTCATTCATTTCTTCTCTGTA 760
 Db 15735 TATTTTATATATTTTTTCTTTTTTTTGTGTTTATGATATATTTTTTTTTTTTA 15676
 QY 761 TTCCTTCTTATCATCTTCATTTGCCAATTTTTCATTTGACATTCATCGTTATG 820
 Db 15675 ATGTTTTTTTTTCTTC-TTTTGTGTTTTATTTTTTTTATATCATTTTTTTTATATA 15617

```
QY 821 TATGATTTTGAACATTTAATGTGA 845
Db 15616 AAAATTTTAAATTTTAAATTTTGA 15592

RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14
Query Match 1.2%; Score 62.6; DB 1; Length 7218;
Best Local Similarity 5.0%; Pred. No. 0.00076;
Matches 20; Conservative 224; Mismatches 153; Indels 0; Gaps 0;

QY 1826 AATTATGGATACAGAGTAATTTGTATATATGATGTGTAGGAGCTAAATAGCAACA 1885
Db 1461 AGTTAAGAGATAGAAGAAATTTGTCACRTRRRRRRRRRRRRRRRRRRRRRRRRR 1402

QY 1886 ATACGTAAGTAATCGAAATCATAAACGTTTAAGGAAGAGGTTTACTAAGT 1945
Db 1401 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1342

QY 1946 CTCGTAATCTCATGTAGCTTGTGTCATGATGATCAGATACAGATAAAAGTTGTAAAGAT 2005
Db 1341 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1282

QY 2006 GACAAAGAAAATGAGAAGTAGGCGGAAGTACGAGTTGGTCGACAAATAGGTGAAGAAC 2065
Db 1281 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1222

us-09-824-735-1.rni
QY 2066 CTTTGCTAAGTTAAGTTTCGAGGAACACAGACACATGCTGTAATAATGTCATCAAAAT 2125
Db 1221 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1162

QY 2126 TATGGCTAAGATACATACTTAAGAACAGAAATGGTTGATCAGGTATGTTCTGATTTT 2185
Db 1161 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1102

QY 2186 TTTTACATGGAACCTAAGTTGTTGCGTCAATGAT 2222
Db 1101 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRAT 1065

RESULT 7
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellum
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
US-07-867-106-2
Query Match 1.2%; Score 60.6; DB 1; Length 5852;
Best Local Similarity 50.7%; Pred. No. 0.0019;
Matches 170; Conservative 0; Mismatches 164; Indels 1; Gaps 1;

QY 1223 TTTTGTGACCTGCTTGACTTAATGAAGTTAAATTTGTTAGTTAAATAGAAAT 1282
```


Qy 1485 AAGTTTAAATCACCACCGATAATTGACCCGACGACTAATTCAGGTCGT 1536
Db 11109 TTATTAATTAATAATTTAAATATTTCTTAGCTATTAAAAAAGGCAT 11058

Search completed: October 21, 2003, 09:34:02
Job time : 269 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2003, 00:09:07 ; Search time 1217 Seconds
(without alignments)
11409.950 Million cell updates/sec

Title: US-09-824-735-1

Perfect score: 5144

Sequence: 1 tttttgcccggaaatctg.....acgtttgatcccaacttaa 5144

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5144	100.0	5144	24	Gene encoding Arab
2	414.6	8.1	2000	24	Arabidopsis thalia
3	401.4	7.8	2000	24	Arabidopsis thalia
4	371.2	7.2	2000	24	Arabidopsis thalia
5	234	4.5	1082138	21	Arabidopsis thalia
6	206.8	4.0	1377	24	cDNA encoding Arab
7	183.2	3.6	1341	21	Arabidopsis thalia
8	163.6	3.2	163319	21	Arabidopsis thalia

C	9	152.4	3.0	2000	24	ABZ17294	Arabidopsis thalia
C	10	115.8	2.3	1502	24	AA148584	A thaliana WRKY64
C	11	112.4	2.2	1082138	21	AAF22305	Arabidopsis thalia
	12	105	2.0	8056	25	ABZ10246	Haematopoietic cel
	13	97.6	1.9	404	25	ABX21874	Human GDP-mannose
	14	97.2	1.9	397	25	ABX62736	Arabidopsis thalia
	15	97.2	1.9	1338	21	AAC42994	Arabidopsis thalia
	16	97.2	1.9	1590	21	AAC36536	Arabidopsis thalia
	17	93.8	1.8	8056	25	ABZ10100	Haematopoietic cel
	18	92.2	1.8	364	21	AAA31238	Plant microsateili
	19	91.2	1.8	845	21	AAC46643	Zea mays DNA fragm
C	20	88.4	1.7	8056	25	ABZ10246	Haematopoietic cel
	21	88	1.7	742	24	ABQ65474	Arabidopsis thalia
	22	84	1.6	1869	21	AAA52775	Soybean putative c
	23	84	1.6	1869	25	ACC00752	Glycine max oil tr
	24	83.8	1.6	1707	21	AAA52767	Corn putative cata
	25	83.8	1.6	1707	25	ACC00742	Zea mays oil trait
C	26	83.6	1.6	8056	25	ABZ10100	Haematopoietic cel
	27	83.2	1.6	465	21	AAC38241	Zea mays DNA fragm
	28	83.2	1.6	1073	21	AAC51017	Arabidopsis thalia
	29	83.2	1.6	1622	21	AAC44238	Arabidopsis thalia
C	30	83.2	1.6	103929	21	AAF22287	BAC containing rep
	31	82.2	1.6	1326	24	ABZ12267	Arabidopsis thalia
	32	81.8	1.6	1548	24	ABL34155	Human immune syste
	33	81.2	1.6	7814	22	AAS46530	Tumour suppressor
	34	80.4	1.6	1320	21	AAC46102	Arabidopsis thalia
	35	80.4	1.6	15387	24	ABL32184	Human immune syste
	36	79.2	1.5	299	25	ABX88968	Corn ear-derived p
	37	78.4	1.5	8067	24	ABN80225	Human chemically m
	38	77.2	1.5	512	21	AAA52778	Corn putative cata
	39	76.6	1.5	7306	24	ABL33637	Human immune syste
	40	75.6	1.5	469	21	AAC43576	Zea mays DNA fragm
	41	75.2	1.5	1326	24	ABZ14508	Arabidopsis thalia
	42	74.8	1.5	1198	21	AAC49879	Arabidopsis thalia
	43	74.6	1.5	1470	24	ABZ14173	Arabidopsis thalia
	44	74.2	1.4	5893	24	ABL32859	Human immune syste
	45	74	1.4	83391	24	ABQ67094	Human angiogenesis

ALIGNMENTS

RESULT 1
ID ABK91073 standard; DNA; 5144 BP.

XX ABK91073;

XX 15-NOV-2002 (first entry)

DE Gene encoding Arabidopsis thaliana SOS2 serine/threonine protein kinase.

XX Salt overly sensitive 7; SOS2; serine/threonine protein kinase;

KW salt tolerance; agricultural crop; rice; corn; wheat; cotton;

KW peanut; soybean; plant protectant; plant; transgenic; gene; ds.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

FT CDS 2004..5133

FT /tag= a

FT /product= "SOS2"

FT exon 2004..2168

FT /tag= b

FT /number= 1

FT intron 2169..2254

FT /tag= c

FT /number= 1

FT exon 2255..2317

FT /tag= d

FT /number= 2

FT intron 2318..2415

FT /tag= e

FT exon /number= 2
PT 2416..2487
FT /tag= f
FT /number= 3
FT intron 2488..2572
FT /tag= g
FT /number= 3
FT exon 2573..2680
FT /tag= h
FT /number= 4
FT intron 2681..2777
FT /tag= i
FT /number= 4
FT exon 2778..2852
FT /tag= j
FT /number= 5
FT intron 2853..2920
FT /tag= k
FT /number= 5
FT exon 2921..3100
FT /tag= l
FT /number= 6
FT intron 3101..3224
FT /tag= m
FT /number= 6
FT exon 3225..3314
FT /tag= n
FT /number= 7
FT intron 3315..3417
FT /tag= o
FT /number= 7
FT exon 3418..3540
FT /tag= p
FT /number= 8
FT intron 3541..4051
FT /tag= q
FT /number= 8
FT exon 4052..4171
FT /tag= r
FT /number= 9
FT intron 4172..4399
FT /tag= s
FT /number= 9
FT exon 4400..4516
FT /tag= t
FT intron 4517..4734
FT /tag= u
FT /number= 10
FT exon 4735..4791
FT /tag= v
FT /number= 11
FT intron 4792..4878
FT /tag= w
FT /number= 11
FT exon 4879..4953
FT /tag= x
FT /number= 12
FT intron 4954..5037
FT /tag= y
FT /number= 12
FT exon 5038..5130
FT /tag= z
FT /number= 13
XX
PN US2002095032-A1.
XX
PD 18-JUL-2002.
XX
PD 04-APR-2001; 2001US-0824735.
XX
PR 04-APR-2000; 2000US-194649P.
XX

PA (UYAR-) UNIV ARIZONA.
XX
PI Zhu J, Liu J, Ishitani M, Kim C, Halfter U;
XX WPI; 2002-665801/71.
DR P-PSDB; ABG31356.
XX
XX New SOS2 protein kinase polypeptide and nucleic acids encoding the
PT polypeptide, useful for increasing salt tolerance to plants, especially
PT to agricultural crops -
XX
XX Claim 3; Page 9-12; 28pp; English.
XX
CC The present invention relates to the isolation of an Arabidopsis
CC thaliana gene encoding salt overly sensitive 7 (SOS2) serine/threonine
CC protein kinase. The polynucleotide sequence encoding SOS2 can be used
CC to increase salt tolerance in plants, particularly agricultural crops
CC such as rice, corn, wheat, cotton, peanut and soybean. The
CC polynucleotide sequence encoding SOS2 can also be used to generate
CC transgenic plants, as hybridization probes for RNA, cDNA and DNA to
CC isolate those cDNAs or genes which exhibit a high degree of similarity
CC to the sequence of the SOS2 gene, and as PCR primers for the production
CC of DNA which encodes an enzyme having serine/threonine protein kinase
CC activity. The present sequence encodes Arabidopsis thaliana SOS2
CC serine/threonine protein kinase.
XX
SQ Sequence 5144 BP; 1523 A; 851 C; 943 G; 1827 T; 0 other;

Query Match 100.0%; Score 5144; DB 24; Length 5144;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTGGCGGAAAATCTCGGTTTACGTTTGGCGGAAAATCTCGTTTACGTTT 60
Db 1 TTTTGGCGGAAAATCTCGGTTTACGTTTGGCGGAAAATCTCGTTTACGTTT 60

QY 61 TGGCGGAAAATCTCGGTTTACGTTTGGCGGAAAATCTCGGTTTACGTTT 120
Db 61 TGGCGGAAAATCTCGGTTTACGTTTGGCGGAAAATCTCGGTTTACGTTT 120

QY 121 CGGGAATAATCAGGATTTATGTTTTTGGTGAATAATTAACGATTTTCTCAAT 180
Db 121 CGGGAATAATCAGGATTTATGTTTTTGGTGAATAATTAACGATTTTCTCAAT 180

QY 181 TTCATTGCTTGATATTTAAGAAAATTTGAAAATAATTAATTTTAAATTTGGTTTGA 240
Db 181 TTCATTGCTTGATATTTAAGAAAATTTGAAAATAATTAATTTTAAATTTGGTTTGA 240

QY 241 TGTGTTGGTTAAACCTAAATTCGCATTTGGTTTAGAGATTTTAGTTGGTTTATCAATT 300
Db 241 TGTGTTGGTTAAACCTAAATTCGCATTTGGTTTAGAGATTTTAGTTGGTTTATCAATT 300

QY 301 TACAAAATTTGATGGTTTAAATGGATAAACCAATGGAACCAATTAACCACTAAC 360
Db 301 TACAAAATTTGATGGTTTAAATGGATAAACCAATGGAACCAATTAACCACTAAC 360

QY 361 TCATTTTACTCATCAACCAATTCATCACTCAATTCATCACTCAATTCATTCAG 420
Db 361 TCATTTTACTCATCAACCAATTCATCACTCAATTCATCACTCAATTCATTCAG 420

QY 421 TCAAAAATTTCAACTCATTAGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGAGTT 480
Db 421 TCAAAAATTTCAACTCATTAGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGAGTT 480

QY 481 TTGACCCATTTTGACACCCCTACATATGATCATAAGTTTAAATCAAAAATTTACTATTGA 540
Db 481 TTGACCCATTTTGACACCCCTACATATGATCATAAGTTTAAATCAAAAATTTACTATTGA 540

QY 541 TAACTTTTAAACGGAATTTGTTTGAAGTTTCATTTACGTTTATTTATATAAAAAACATA 600
Db 541 TAACTTTTAAACGGAATTTGTTTGAAGTTTCATTTACGTTTATTTATATAAAAAACATA 600

QY 601 ATGCAAAAGTACTAATGTATAGTTATTTTATTTTAAATGCAAAAATATTACTGTAA 660

Db 601 ATGCAAAAGTACTAATGATAGTATATTTTATTTTAAATGCAAAAATATTACGTAA 660
Qy 661 TACATTTTCACTCTATCAATATTTTCTTATATCAATAGCCACCAACCTACAT 720
Db 661 TACATTTTCACTCTATCAATATTTTCTTATATCAATAGCCACCAACCTACAT 720
Qy 721 TCAATATTTCTTCCACATTCATATTTTCTTGTCTTGTATTCCTTCTTATCAT 780
Db 721 TCAATATTTCTTCCACATTCATATTTTCTTGTCTTGTATTCCTTCTTATCAT 780
Qy 781 TGCCAAATTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Db 781 TGCCAAATTTTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Qy 841 TGTGAATTTTCAACCGACTATAGGAGTCTCAAGAGTTTTCAAAATTTTAAAAAGATCT 900
Db 841 TGTGAATTTTCAACCGACTATAGGAGTCTCAAGAGTTTTCAAAATTTTAAAAAGATCT 900
Qy 901 TGAACCTTTTAAAGCAAAATCCGACCGACTATAGAAACCTATAGAACTTTTAAAAAGATCT 960
Db 901 TGAACCTTTTAAAGCAAAATCCGACCGACTATAGAAACCTATAGAACTTTTAAAAAGATCT 960
Qy 961 AGCTCTTCTGTCGCGCGGAAATCTAGGCAATAGAAAGCTCTTCAACATCACTAGTAT 1020
Db 961 AGCTCTTCTGTCGCGCGGAAATCTAGGCAATAGAAAGCTCTTCAACATCACTAGTAT 1020
Qy 1021 ATTGACCGTGACCACTTTTGTGACCAATGCTTTGTGAATGAACCGTCGATAAACCGTGT 1080
Db 1021 ATTGACCGTGACCACTTTTGTGACCAATGCTTTGTGAATGAACCGTCGATAAACCGTGT 1080
Qy 1081 ATCACTTACCGCAAAATTTTCCCTAGTGTGTTTGTGTTTCCAACTCTCAGAACTCTTATCGA 1140
Db 1081 ATCACTTACCGCAAAATTTTCCCTAGTGTGTTTGTGTTTCCAACTCTCAGAACTCTTATCGA 1140
Qy 1141 ACTTTTATATATACACTTTGTAGCAATGGAAGTATCTTTGTATGCTTTGTCTTAACT 1200
Db 1141 ACTTTTATATATACACTTTGTAGCAATGGAAGTATCTTTGTATGCTTTGTCTTAACT 1200
Qy 1201 TAGACATCTTGTCTCTGTTTGTGTTTGTGAACTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 1260
Db 1201 TAGACATCTTGTCTCTGTTTGTGTTTGTGAACTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 1260
Qy 1261 TTGTAGTAAATAATAGAAATTTTACTAAATTTGAGTTTGCATTAATCATAGTCTAGATAAT 1320
Db 1261 TTGTAGTAAATAATAGAAATTTTACTAAATTTGAGTTTGCATTAATCATAGTCTAGATAAT 1320
Qy 1321 TTGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1380
Db 1321 TTGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1380
Qy 1381 CTAAATGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440
Db 1381 CTAAATGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1440
Qy 1441 ACAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 1500
Db 1441 ACAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 1500
Qy 1501 AACGATAATTTGACCGCAACGACTAATTCAGGTGCTATACGGGTACAGTCAAGTCAAGTCAAGTCAAG 1560
Db 1501 AACGATAATTTGACCGCAACGACTAATTCAGGTGCTATACGGGTACAGTCAAGTCAAGTCAAGTCAAG 1560
Qy 1561 ACCGAAATGCTCAACCGATCTGACCGAGCTTCCGATCGACGATTCACGCTTTTCGCGA 1620
Db 1561 ACCGAAATGCTCAACCGATCTGACCGAGCTTCCGATCGACGATTCACGCTTTTCGCGA 1620
Qy 1621 ATCGCATCACGAGCTTCT 1680
Db 1621 ATCGCATCACGAGCTTCT 1680
Qy 1681 CAATTTCTGATTTTCAAGATTTCTTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740

Db 1681 CAATTTCTGATTTTCAAGATTTCTTTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1740
Qy 1741 ATTTTATATATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 1800
Db 1741 ATTTTATATATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 1800
Qy 1801 ATGTTGATTTATGCTGCGAAATCAAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1860
Db 1801 ATGTTGATTTATGCTGCGAAATCAAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1860
Qy 1861 GTGTAGGAGCTTAAATTTAGGCAACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1920
Db 1861 GTGTAGGAGCTTAAATTTAGGCAACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1920
Qy 1921 TTAAGGAAAGAGGTTTCTTAACTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1980
Db 1921 TTAAGGAAAGAGGTTTCTTAACTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1980
Qy 1981 GATCAGATATAAGTTTGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
Db 1981 GATCAGATATAAGTTTGTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2040
Qy 2041 TTGTTGCGCACAAATAGGTGAAGAACCTTTTCTAAGGTTTAAAGTTTTCGAGGAAACACAGACA 2100
Db 2041 TTGTTGCGCACAAATAGGTGAAGAACCTTTTCTAAGGTTTAAAGTTTTCGAGGAAACACAGACA 2100
Qy 2101 CTGTTGATATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2160
Db 2101 CTGTTGATATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2160
Qy 2161 TTGATCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220
Db 2161 TTGATCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220
Qy 2221 ATGATCTTTGATTTTCTGTTTAAAGCTCTTTTACAGATAAAGAGAGATATCTTATAATGAA 2280
Db 2221 ATGATCTTTGATTTTCTGTTTAAAGCTCTTTTACAGATAAAGAGAGATATCTTATAATGAA 2280
Qy 2281 GATTTTCTGTCACCGCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340
Db 2281 GATTTTCTGTCACCGCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340
Qy 2341 TCTCGAAATTTTATCTCTGAAAGTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2400
Db 2341 TCTCGAAATTTTATCTCTGAAAGTTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2400
Qy 2401 ATTTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2460
Db 2401 ATTTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2460
Qy 2461 ACAGGAGGAGGATCTTTTGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2520
Db 2461 ACAGGAGGAGGATCTTTTGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2520
Qy 2521 TTAGTTAAACCTCTCTACTTAAATTTTAAATATGATGATGATGATGATGATGATGATGATGATGATG 2580
Db 2521 TTAGTTAAACCTCTCTACTTAAATTTTAAATATGATGATGATGATGATGATGATGATGATGATGATG 2580
Qy 2581 AGGAGGCTTCAAGAAAGTGTCTCGGAATCTTTTCAACAGCTTGTAGATGATGATGATGATGATGATG 2640
Db 2581 AGGAGGCTTCAAGAAAGTGTCTCGGAATCTTTTCAACAGCTTGTAGATGATGATGATGATGATGATG 2640
Qy 2641 TCAATGTCATCTGCAAGGTTTAAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 2700
Db 2641 TCAATGTCATCTGCAAGGTTTAAACCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 2700
Qy 2701 ACCAATATTTCTCAGAAATCTCTCAGCTGTTGAAATCCAGACTTGTATGATGATGATGATGATGATG 2760
Db 2701 ACCAATATTTCTCAGAAATCTCTCAGCTGTTGAAATCCAGACTTGTATGATGATGATGATGATGATG 2760
Qy 2761 TATGTTATGTTTATCTAGCCAGAAATCTTTTCTGATGATGATGATGATGATGATGATGATGATGATG 2820
Db 2761 TATGTTATGTTTATCTAGCCAGAAATCTTTTCTGATGATGATGATGATGATGATGATGATGATGATG 2820

QY	2821	CGGATTTTCGGACTCAGTGCATTTGCTCTCAGGAGTAAGTGCTCTTATCTCTGCTTTCAGCAG	2888
DB	2821	CGGATTTTCGGACTCAGTGCATTTGCTCTCAGGAGTAAGTGCTCTTATCTCTGCTTTCAGCAG	2889
QY	2881	TCTGCTTACGTGGTCAATTAACCTTGTTATATACTCAATCAGGGAGTAGAACTTCTGCTGTAC	2940
DB	2881	TCTGCTTACGTGGTCAATTAACCTTGTTATATACTCAATCAGGGAGTAGAACTTCTGCTGTAC	2940
QY	2941	CACATGTGGAACTCCGAACTGATGTAGCTCCAGAGGTACTTAGTGGACAGGGTTACGATGG	3000
DB	2941	CACATGTGGAACTCCGAACTGATGTAGCTCCAGAGGTACTTAGTGGACAGGGTTACGATGG	3000
QY	3001	TTCAGCAGCTGATATTTGGTCTTCGGGGGTATTTCTTTTCGTTATATTGCTCGATATTT	3060
DB	3001	TTCAGCAGCTGATATTTGGTCTTCGGGGGTATTTCTTTTCGTTATATTGCTCGATATTT	3060
QY	3061	ACCTTTTTCGAGACGGATCTTCAGAGGTTGTACAGAAAAAGTAAGTAACATATCTTTTCGG	3120
DB	3061	ACCTTTTTCGAGACGGATCTTCAGAGGTTGTACAGAAAAAGTAAGTAACATATCTTTTCGG	3120
QY	3121	GAAGAAATCATGAATTCCTTGTCATGCTTTTGTCAAAACCGTTTATTGATTTTGGTTTTGC	3180
DB	3121	GAAGAAATCATGAATTCCTTGTCATGCTTTTGTCAAAACCGTTTATTGATTTTGGTTTTGC	3180
QY	3181	AAATTCACCCCTTAGTTTGTAGCTTTTACATTTGTTATTTTACAGATAAATGCGACGAGGT	3240
DB	3181	AAATTCACCCCTTAGTTTGTAGCTTTTACATTTGTTATTTTACAGATAAATGCGACGAGGT	3240
QY	3241	TTTCTTGTCCACCGTGTGTTTCCGCAAGAGTGAAAGTTTTTAAATACATAGATATCTTTGACC	3300
DB	3241	TTTCTTGTCCACCGTGTGTTTCCGCAAGAGTGAAAGTTTTTAAATACATAGATATCTTTGACC	3300
QY	3301	CCAAATCCCAAAAAGTGGAGTATTTTGTTTGTCTCTCTAGCTATCAGGTTTGGTGAT	3360
DB	3301	CCAAATCCCAAAAAGTGGAGTATTTTGTTTGTCTCTCTAGCTATCAGGTTTGGTGAT	3360
QY	3361	ATTATAATGTTCTAGTAATATATCTGTTTATCTATTATGTTTCTCAATTAGACGACGT	3420
DB	3361	ATTATAATGTTCTAGTAATATATCTGTTTATCTATTATGTTTCTCAATTAGACGACGT	3420
QY	3421	ATTCAAAATTCAGGAATCAAGAAAGATCCCTTGTTGTCAGATTAATATGTCGCTATACGA	3480
DB	3421	ATTCAAAATTCAGGAATCAAGAAAGATCCCTTGTTGTCAGATTAATATGTCGCTATACGA	3480
QY	3481	GCAAGGAAAGAAAGAAAGTGAATTTGGATGATATTCGTGACGTTTTTGTGAAATGGAG	3540
DB	3481	GCAAGGAAAGAAAGAAAGTGAATTTGGATGATATTCGTGACGTTTTTGTGAAATGGAG	3540
QY	3541	GTGTTGTGCTTCGCTTCATTAATATGCTCTTTGCTGGTCAATTCATTTAAATGTTAAGAT	3600
DB	3541	GTGTTGTGCTTCGCTTCATTAATATGCTCTTTGCTGGTCAATTCATTTAAATGTTAAGAT	3600
QY	3601	CTCTTAGGAACGTTTGGATGACCAAGAAAGTGTGTTGCTACAGGATAGAAACAAATAGT	3660
DB	3601	CTCTTAGGAACGTTTGGATGACCAAGAAAGTGTGTTGCTACAGGATAGAAACAAATAGT	3660
QY	3661	AGGCATGTGTTAGTTACCAAACTGTAACTGCTTTTATTCAAATTCGCCAAACCATAG	3720
DB	3661	AGGCATGTGTTAGTTACCAAACTGTAACTGCTTTTATTCAAATTCGCCAAACCATAG	3720
QY	3721	ACCTTAGGAGACTTAGATTTACAGAGAAATTCCTCTATTCTCGAACCAAAACCTTAGACA	3780
DB	3721	ACCTTAGGAGACTTAGATTTACAGAGAAATTCCTCTATTCTCGAACCAAAACCTTAGACA	3780
QY	3781	AAATCCAGATAATCCCTTAGGCTTAAATACAACTGTTTCATGTACCTATATATATCTCC	3840
DB	3781	AAATCCAGATAATCCCTTAGGCTTAAATACAACTGTTTCATGTACCTATATATATCTCC	3840
QY	3841	TGCTAAGATGCTTAGCTTTGGTATAGCTAGATATATAGATACCGTAATTTCTAAT	3900
DB	3841	TGCTAAGATGCTTAGCTTTGGTATAGCTAGATATATAGATACCGTAATTTCTAAT	3900

[illegible]

Db 4981 AGTTGCATCTACTGTTGCGGATTAGAGTGATATTTTGTGTTATCGAGTTC 5040
 Qy 5041 TACAGAGCTATGTTGGAACCTGGAACACATAATATGAGGGCAACAGAGGAATACCA 5100
 Db 5041 TACAGAGCTATGTTGGAACCTGGAACACATAATATGAGGGCAACAGAGGAATACCA 5100
 Qy 5101 AAGTCAGAGATCTCAGAGCAATCACTGTTTGTATCCCACTTAA 5144
 Db 5101 AAGTCAGAGATCTCAGAGCAATCACTGTTTGTATCCCACTTAA 5144

RESULT 2

ABZ14979
 ID ABZ14979 standard; DNA; 2000 BP.

XX AC ABZ14979;

XX DT 21-JAN-2003 (first entry)

XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 2784.

XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

XX XX Arabidopsis thaliana.

XX PN WO200216655-A2.

XX PD 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26685.

XX PR 24-AUG-2000; 2000US-227866P.

XX PR 26-JAN-2001; 2001US-264647P.

XX PR 22-JUN-2001; 2001US-300111P.

XX PA (SCRI) SCRIPPS RES INST.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Harper JF, Kreps J, Wang X, Zhu T;

XX DR WPI; 2002-304127/34.

XX PT Identifying a stress condition to which a plant cell has been exposed

XX PS Claim 144; SEQ ID NO 2784; 577bp + Sequence Listing; English.

XX CC The invention relates to identifying a stress condition to which a plant

XX CC cell has been exposed, comprising:

XX CC (a) contacting nucleic acid representative of expressed polynucleotides

XX CC in the plant cell with an array or probes representative of the plant

XX CC cell genome; and

XX CC (b) detecting a profile of expressed polynucleotides in the plant cell

XX CC characteristic of a stress response. The method is useful in the

XX CC production of transgenic plants, cells and seeds and in producing plants

XX CC with increased tolerance to abiotic stress. The present sequence is that

XX CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used

XX CC in methods of the invention.

XX CC Note: the sequence data for this patent is not represented in the printed

XX CC specification but is based on sequence information supplied to Derwent by

XX CC the European Patent Office.

XX SQ Sequence 2000 BP; 675 A; 270 C; 353 G; 702 T; 0 other;

Query Match

Best Local Similarity 8.1%; Score 414.6; DB 24; Length 2000;

Matches 489; Conservative 0; Mismatches 59; Indels 9; Gaps 3;

Qy 1 TTTTGGCGGGAATCTCGGGTTTACGTTTGGCGGGAATCTCGGGTTTACGTTT 60

Db 510 TTTTGGTGAGAAATTCGGGTTTACGTTTGGCGGGAATTCGGGTTTACGTTT 569

Qy 61 TGGCGGGAATCTCGGGTTTACGTTTTCGGAGAAAATCAOCCGGTTTACTTTTTTGG 120
 Db 570 TGGCGGGAATCTCGGGTTTACGTTTTCGGAGAAAATCAOCCGGTTTAC-GTTTTTGG 628
 Qy 121 CGGGAATTCACGGATTTATGTTTTCGGTGGAAAAATTCAGAGTTTACTTTTCTCAAT 180
 Db 629 CGGGAATTCACGGGTTTATGTTTTCGGTGGAAAAATTCAGAGTTTATTTTCTCAAT 688
 Qy 181 TCCATTGCTGATATTTAAGAAATTTGGAAAAATTAATTAATTTTAAATTTGGTTAGA 240
 Db 689 TCCATCGCTTGATATTTAAGAAATTTGGAAAAATTAATTAATTTTAAATTTGGTTAGA 748
 Qy 241 TGTGTTGGTTAAACCTAAATTTGGCAATTTGGTATTAGAGATTTTAGTTGGTTTAAATTT 300
 Db 749 TGTGTTGGTTAAACCTAAATTTGGCAATTTGGTATTAGAGATTTTAGTTGGTTTAAATTT 808
 Qy 301 TACAAAATTTGATGGTTAATTTGATTAACCAATGGAACCAATTAACCAATTAACCAATTAAC 360
 Db 809 TACAAAATTTGATGGTTAATTTGATTAACCAATGGAACCAATTAACCAATTAACCAATTAAC 868
 Qy 361 TCAATTTTACTCATCAACCAATTTGACTCATCAACTCATTTGACTCATCAACTCATTTGAG 420
 Db 869 TCAATTTTACTCATCAACCAATTTGACTCATCAACTCATTTGACTCATCAACTCATTTGAG 928
 Qy 421 TCAAAAATTTCAACTCATTTAGGGTTTCATGGTTTCATGGTTTCATGGTTTCATGGTTTCAT 480
 Db 929 TAAAAAATTT-----TAAAGGTTTCATGGTTTCATGGTTTCATGGTTTCATGGTTTCAT 981
 Qy 481 TTGACCCATTTTGACACCCCT-ACATATGATCATATGATTAATTAATTAATTAATTAATTA 539
 Db 982 TTGACCCATTTTGACACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCC 1041
 Qy 540 ATAACCTTTTAAACGGAA 556
 Db 1042 TAGTTTATTTTATTA 1058

RESULT 3

ABZ17522/c

ID ABZ17522 standard; DNA; 2000 BP.

XX AC ABZ17522;

XX DT 21-JAN-2003 (first entry)

XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 5327.

XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

XX OS Arabidopsis thaliana.

XX PN WO200216655-A2.

XX PD 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26685.

XX PR 24-AUG-2000; 2000US-227866P.

XX PR 26-JAN-2001; 2001US-264647P.

XX PR 22-JUN-2001; 2001US-300111P.

XX PA (SCRI) SCRIPPS RES INST.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Harper JF, Kreps J, Wang X, Zhu T;

XX DR WPI; 2002-304127/34.

XX PT Identifying a stress condition to which a plant cell has been exposed

XX PS and producing plants with increased tolerance to these abiotic stresses

PS Claim 144; SEQ ID NO 5327; 577pp + Sequence Listing; English.
 XX The invention relates to identifying a stress condition to which a plant
 CC cell has been exposed, comprising:
 CC (a) contacting nucleic acid representative of expressed polynucleotides
 CC in the plant cell with an array or probes representative of the plant
 CC cell genome; and
 CC (b) detecting a profile of expressed polynucleotides in the plant cell
 CC characteristic of a stress response. The method is useful in the
 CC production of transgenic plants, cells and seeds and in producing plants
 CC with increased tolerance to abiotic stress. The present sequence is that
 CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
 CC in methods of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office.
 XX SQ Sequence 2000 BP; 681 A; 386 C; 275 G; 658 T; 0 other;
 Query Match 7.8%; Score 401.4; DB 24; Length 2000;
 Best Local Similarity 89.4%; Pred. No. 1.7e-58;
 Matches 455; Conservative 0; Mismatches 51; Indels 3; Gaps 2;
 QY 2 TTTTGGCGGAAATCTCGGTTTACGTTTGGCGGAAATCTCGGTTTACGTTT 61
 Db |||||
 QY 1187 TTTTGGCGGAAATCTCGGTTTACGTTTGGCGGAAATCTCGGTTTACGTTT 1128
 Db |||||
 QY 62 GCGCGGAAATCTCGGTTTACGTTTGGCGGAAATCTCGGTTTACGTTT 121
 Db |||||
 QY 1127 GCGCGGAAATCTCGGTTTACGTTTGGCGGAAATCTCGGTTTACGTTT 1069
 Db |||||
 QY 122 GCGGAAATCTCGGTTTACGTTTGGCGGAAATCTCGGTTTACGTTT 181
 Db |||||
 QY 1068 GCGGAAATCTCGGTTTACGTTTGGCGGAAATCTCGGTTTACGTTT 1011
 Db |||||
 QY 182 TCATTGCTTGATATTAAGAAATTTGAAATATTAATTTTAAATTTAGAT 241
 Db |||||
 QY 1010 TCATGCTTGATATTAAGAAATTTGAAATATTAATTTTAAATTTAGAT 951
 Db |||||
 QY 242 GGTGTTGTTAACTTAATTTGCAATTTGTTAGATTTAGTTGGTTTATTCATTT 301
 Db |||||
 QY 950 GGTGTTGTTAACTTAATTTGCAATTTGTTAGATTTAGTTGGTTTATTCATTT 891
 Db |||||
 QY 302 ACAAAATTTGATGGTTAATTTGATAAACCATTGAAACCATTAACCACTCACT 361
 Db |||||
 QY 890 ACAAAATTTGATGGTTAATTTGATAAACCATTGAAACCATTAACCACTCACT 831
 Db |||||
 QY 362 CATTTTACTCATCAACCAATTTGACTCATCACTCATTTGACTCATCACTCATTTGAGT 421
 Db |||||
 QY 830 CATTTTACTCATCAACCAATTTGACTCATCACTCATTTGACTCATCACTCATTTGAGT 771
 Db |||||
 QY 422 CAATAATTTCACTCATTTAGGTTTCACTGGTTGAGTTGAGTTGAGTTGAGTTGAGTT 481
 Db |||||
 QY 770 CAATAATTTCACTCATTTAGGTTTCACTGGTTGAGTTGAGTTGAGTTGAGTTGAGTT 711
 Db |||||
 QY 482 TGACCCATTTTGACACCCCTCATATGAT 510
 Db |||||
 QY 710 TAAACCATTTTGACATCTCTACCTCGGT 682
 Db |||||
 RESULT 4
 ABZ16852
 ID ABZ16852 standard; DNA; 2000 BP.
 XX AC ABZ16852;
 XX DT 21-JAN-2003 (first entry)
 XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 4657.
 XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
 XX OS Arabidopsis thaliana.

XX WO200216655-A2.
 XX 28-FEB-2002.
 XX 24-AUG-2001; 2001WO-US26685.
 XX 24-AUG-2000; 2000US-227866P.
 XX 26-JAN-2001; 2001US-264647P.
 XX 22-JUN-2001; 2001US-300111P.
 XX (SCRI) SCRIPPS RES INST.
 XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX Harper JF, Kreps J, Wang X, Zhu T;
 XX WPI; 2002-304127/34.
 XX Identifying a stress condition to which a plant cell has been exposed
 XX and producing plants with increased tolerance to these abiotic stresses
 XX Claim 144; SEQ ID NO 4657; 577pp + Sequence Listing; English.
 XX The invention relates to identifying a stress condition to which a plant
 XX cell has been exposed, comprising:
 XX (a) contacting nucleic acid representative of expressed polynucleotides
 XX in the plant cell with an array or probes representative of the plant
 XX cell genome; and
 XX (b) detecting a profile of expressed polynucleotides in the plant cell
 XX characteristic of a stress response. The method is useful in the
 XX production of transgenic plants, cells and seeds and in producing plants
 XX with increased tolerance to abiotic stress. The present sequence is that
 XX of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
 XX in methods of the invention.
 XX Note: The sequence data for this patent is not represented in the printed
 XX specification but is based on sequence information supplied to Derwent by
 XX the European Patent Office.
 XX SQ Sequence 2000 BP; 708 A; 255 C; 280 G; 757 T; 0 other;
 Query Match 7.2%; Score 371.2; DB 24; Length 2000;
 Best Local Similarity 95.3%; Pred. No. 2e-53;
 Matches 404; Conservative 0; Mismatches 18; Indels 2; Gaps 2;
 QY 47 CGTGTTCAGTTTGGCGGAAATCTCGGTTTACGTTTTCGCGGAAATCTCGG 106
 Db |||||
 QY 748 CGGTTTACGTTTGGCGGAAATCTCGGTTTACGTTTTCGCGGAAATCTCGG 807
 QY 107 TTTTACCTTTTGGCGGAAATCTCGGTTTATGTTTGGTGGAAATTTACGATT 166
 Db |||||
 QY 808 TTTTAC-GTTTGGCGGAAATCTCGGTTTACG-TTTTGGCGGAAATTTACGGTT 865
 QY 167 TACTTTTCTCAATTTCAATTTGCTGTATATTTAAGAAATTTGGAATTTATTTAT 226
 Db |||||
 QY 866 TACTTTTCTCAATTTCAATTTGCTGTATATTTAAGAAATTTGGAATTTATTTAT 925
 QY 227 TAAATTTGTTTGTAGTGTGGTTAACTAACTAAATTTGGCATTGGTTAGATTG 286
 Db |||||
 QY 926 TAAATTTGTTTGTAGTGTGGTTAGCTTAAATTTGGCATTGGTTAGATTGTTAG 985
 QY 287 GTTTTATTCATTTTACAAATTTGATGGTTTAAATTTGGATTAACCATGAACCACTTAAC 346
 Db |||||
 QY 986 GTTTTATTCATTTTACAAATTTGATGGTTTAAATTTGGATTAACCATGAACCACTTAAC 1045
 QY 347 CATTAACCACTTAACCATTTTACTCATCAACCAATTTGACTCATCACTCATTTGACTCA 406
 Db |||||
 QY 1046 CATTAACCACTTAACCATTTTACTCATCAACCAATTTGACTCATCACTCATTTGACTCA 1105
 QY 407 TCAACTCATTTGAGTCAAAATTTTCAACTCATTTGAGTTTCACTGGTTTGGTTGAGTTGAG 466
 Db |||||
 QY 1106 TCAACTCATTTGAGTCAAAATTTTCAACTCATTTGAGTTTCACTGGTTTGGTTGAGTTGAG 1165

```

QY      467 TTGA 470
      |||||
Db      1166 TTGA 1169

RESULT 5
AAF22305
ID      AAF22305 standard; DNA; 1082138 BP.
XX
AC      AAF22305;
XX
XX      20-MAR-2001 (first entry)
XX
XX      Arabidopsis thaliana chromosome 4 centromere.
XX
KW      Centromere; microsome; vector; ds.
XX
XX      Arabidopsis thaliana.
XX
XX      WO200055325-A2.
XX
XX      21-SEP-2000.
XX
XX      17-MAR-2000; 2000WO-US07392.
XX
XX      18-MAR-1999; 99US-0125219.
XX
XX      01-APR-1999; 99US-0127409.
XX
XX      18-MAY-1999; 99US-0134770.
XX
XX      13-SEP-1999; 99US-0153584.
XX
XX      17-SEP-1999; 99US-0154603.
XX
XX      (UVCN-) UNIV CHICAGO.
XX
XX      Preuss D, Copenhaver G, Keith K;
XX
XX      WPI; 2000-587529/55.
XX
XX      Recombinant DNA construct comprising a plant centromere, useful for
XX      producing stably inherited microchromosomes which can serve as vectors for
XX      the construction of transgenic plant and animal cells -
XX
XX      Claim 68; Page 977-1388; 1449pp; English.
XX
XX      The present invention relates to a recombinant DNA construct of a plant
XX      (Arabidopsis thaliana) centromere. The constructs are useful for
XX      producing stably inherited microchromosomes which can serve as vectors for
XX      the construction of transgenic plant and animal cells expressing
XX      selected proteins such as hormones, enzymes, interleukins, clotting
XX      factors, cytokines, antibodies, and growth factors.
XX
XX      Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;
XX
XX      Query Match      4.5%; Score 234; DB 21; Length 1082138;
XX      Best Local Similarity 82.8%; Pred. No. 3.7e-30;
XX      Matches 308; Conservative 0; Mismatches 50; Indels 14; Gaps 3;
XX
QY      218 TAATTTTATTAATGGTTTAGATGCTGTGGTAAACCTAAATGGCATGGTTTACAGA 277
      |||||
Db      822201 TGATTTTATTAATGGTTTAGATGCTGTGGTAAAT-----ATTGGTTTACAGA 822250
      |||||

QY      278 TTTTAGTGGTTTATTCATTTTACAAAATTTGATGGGTTAATTTGATATAACCATGGAA 337
      |||||
Db      822251 TTTTAGTGGTTTATTAATTCATTTTACAAAATTTGATGGGTTAATTTGGTTAAACCATGAA 822310
      |||||

QY      338 ACCATTAA-CCATTACACCTACTCATTTTACTCATCAACCAATGACTCATCACTC 396
      |||||
Db      822311 ACCATTAAACCCATTACACCACTCATTTTACTCATCAACCAATGACTCATCACTC 822370
      |||||

QY      397 ATTGACTCATCACTCATTTGACTCAAAATTTCACTCATTTAGGTTTCATGGTTGAG 456
      |||||
Db      822371 ATTGACCCATCACTCATTTGACTCAAAATTTCACTCATTTAGGTTTCATGGTTGAG 822430
      |||||

QY      457 TTGAGTTGAGTTGACCCATCAATTTTGACCCATTTTGACCCATTTTGACCCATTTTGACCAT 513
      |||||

```

```

Db      822431 TTGAGTTGAGTTGACCCATCAATTTTGACCCATTTTGACCCATTTTGACCCATTTTGACCCAT 822490
      |||||
QY      514 AAGTTAATATCAAAAATTTACTATTGATACCTTTTAAACGGAATGTTTGTAAAGTTTCA 573
      |||||
Db      822491 AATTAATAACTCTCTACAGTTACATATATTTTTCACAAAACATTTTGAAAAATAAT 822550
      |||||
QY      574 TTTACGTTATTT 585
      |||||
Db      822551 TGTATCAATTT 822562
      |||||

RESULT 6
ABK91084
ID      ABK91084 standard; cDNA; 1377 BP.
XX
XX      AC      ABK91084;
XX
XX      15-NOV-2002 (first entry)
XX
XX      cDNA encoding Arabidopsis thaliana SOS2 serine/threonine protein kinase.
XX
XX      Salt overly sensitive 7; SOS2; serine/threonine protein kinase;
XX      salt tolerance; agricultural crop; rice; corn; wheat; cotton;
XX      peanut; soybean; plant protectant; plant; transgenic; gene; ss.
XX
XX      Arabidopsis thaliana.
XX
XX      Key      Location/Qualifiers
XX      CDS      26..1366
XX              /tag= a
XX              /product= "SOS2"
XX
XX      US2002095032-A1.
XX
XX      18-JUL-2002.
XX
XX      04-APR-2001; 2001US-0824735.
XX
XX      04-APR-2000; 2000US-194649P.
XX
XX      (UYAR-) UNIV ARIZONA.
XX
XX      Zhu J, Liu J, Ishitani M, Kim C, Halfter U;
XX
XX      WPI; 2002-665801/71.
XX      P-PSDB; ABG31356.
XX
XX      New SOS2 protein kinase polypeptide and nucleic acids encoding the
XX      polypeptide, useful for increasing salt tolerance to plants, especially
XX      to agricultural crops -
XX
XX      Disclosure; Fig 2; 28pp; English.
XX
XX      The present invention relates to the isolation of an Arabidopsis
XX      thaliana gene encoding salt overly sensitive 7 (SOS2) serine/threonine
XX      protein kinase. The polynucleotide sequence encoding SOS2 can be used
XX      to increase salt tolerance in plants, particularly agricultural crops
XX      such as rice, corn, wheat, cotton, peanut and soybean. The
XX      polynucleotide sequence encoding SOS2 can also be used to generate
XX      transgenic plants, as hybridization probes for RNA, cDNA and DNA to
XX      isolate those cDNAs or genes which exhibit a high degree of similarity
XX      to the sequence of the SOS2 gene, and as PCR primers for the production
XX      of DNA which encodes an enzyme having serine/threonine protein kinase
XX      activity. The present sequence encodes Arabidopsis thaliana SOS2
XX      serine/threonine protein kinase.
XX
XX      Sequence 1377 BP; 434 A; 221 C; 341 G; 381 T; 0 other;
XX
XX      Query Match      4.0%; Score 206.8; DB 24; Length 1377;
XX      Best Local Similarity 61.3%; Pred. No. 8.4e-26;
XX      Matches 692; Conservative 0; Mismatches 2; Indels 434; Gaps 5;

```

QY 1979 TGGATCAGATAAAAGTTTGTAAAGATGACAAAGAAATGAGAGAGTGGCAAGTACGA 2038
Db 1 TGGATCAGATAAAAGTTTGTAAAGATGACAAAGAAATGAGAGAGTGGCAAGTACGA 60
QY 2039 GGTGGTCGCACATAGTAGGAAGAACCTTTGCTAAAGTTTAAAGTTTGCAGGAACACAGA 2098
Db 61 GGTGGTCGCACATAGTAGGAAGAACCTTTGCTAAAGTTTAAAGTTTGCAGGAACACAGA 120
QY 2099 CACTGGTGATATGTAGCCATCAAAATTTATGGCTAAGAGTACAATCTTAAGAACAAGAT 2158
Db 121 CACTGGTGATATGTAGCCATCAAAATTTATGGCTAAGAGTACAATCTTAAGAACAAGAT 180
QY 2159 GGTTCATCAGGTATGTTCTGCAATGTTTTTACATGGAACCTAAAGTTTGTGCGTCAATG 2218
Db 181 GGTTCAT- 187
QY 2219 GTATGATCTTTGATTTGCTTTTAAAGCTCTTTTACAGATAAAAGAGAGATATCTATAATG 2278
Db 188 -----CAGATAAAAGAGAGATATCTATAATG 214
QY 2279 AAGATTGTTCTGCACCGAACATAGTAGAGTTGTATGAGGTATGTTTGTGTTCCATG 2338
Db 215 AAGATTGTTCTGCACCGAACATAGTAGAGTTGTATG----- 251
QY 2339 CATCTGCGAAATTTATCTCTGAAGTGTTTTTCATCATTTCTTCTGTTGTTTTTTG 2398
Db 252 ----- 251
QY 2399 TGATTTCCGATGTAGTGTGTGGAGTCTCTCGAATAATATATATAGTTTGGAGTTG 2458
Db 252 -----AGGTGTGGCGAGTCTCTCGAATAATATATATAGTTTGGAGTTG 296
QY 2459 TGACAGGAGGAGAGCTCTTTGATAGAATTTGACGGAACCTCCATCTACTGTAGGACGGTC 2518
Db 297 TGACAGGAGGAGAGCTCTTTGATAGAAT----- 325
QY 2519 CATTAGTTAAACCTCTCTACTTAAATTTTAAATATATGAATCTTTTCATGCAAGTTTCAT 2578
Db 326 -----GTTTCAT 331
QY 2579 AAGGGAGGCTTGAGAAAGTAGTCTCGGAATATCTTTCAACAGCTTGTAGATGCTGTT 2638
Db 332 AAGGGAGGCTTGAGAAAGTAGTCTCGGAATATCTTTCAACAGCTTGTAGATGCTGTT 391
QY 2639 GCTCATGTGCTGCAAGGCTGTTTACACCGTGACCTAAAGGTAAAGACGCTGTTTGT 2698
Db 392 GCTCATGTGCTGCAAGGCTGTTTACACCGTGACCTAA----- 431
QY 2699 TTACCAATATCTCTAGATATCTCACTGCGTTGCAATCCAGACTTGATATTTTGTGTC 2758
Db 432 ----- 431
QY 2759 GCTATGTTATGTTATCTAGCCAGAAATCTTTTACTCGATACAAATCGAAATCTGAAGGT 2818
Db 432 -----AGCCAGAAATCTTTTACTCGATACAAATCGAAATCTGAAGGT 474
QY 2819 TTCGGATTTGGGACTCAGTGCATTCCTCAGGAAGTAAGTCTTATCTCTCTCTTCAGC 2878
Db 475 TTCGGATTTGGGACTCAGTGCATTCCTCAGGA----- 508
QY 2879 AGTCTGCTTACGTGGTCAATTAATGTTTATATCTCAATCAGGGAGTAGAATCTTCGCT 2938
Db 509 -----GGAGTAGAATCTTCGCT 526
QY 2939 ACCACATGGAACCTCGACTATGTAGCTCCAGAGTACTTAGTGGACAGGGTTACGAT 2998
Db 527 ACCACATGGAACCTCGACTATGTAGCTCCAGAGTACTTAGTGGACAGGGTTACGAT 586
QY 2999 GGTTCAGCAGCTGATATTTGGTCTTGGCGGGTTATTTCTTTCGTTATATTTGCTGATAT 3058
Db 587 GGTTCAGCAGCTGATATTTGGTCTTGGCGGGTTATTTCTTTCGTTATATTTGCTGATAT 646
QY 3059 TTACCTTTTCCGAGACGGATCTTCCAGGGTTGTACAGAAAAAGTAAGT 3106

647 TTACCTTTTCCGAGACGGATCTTCCAGGGTTGTACAGAAAAATAAT 694

RESULT 7

AAC50382
ID AAC50382 standard; DNA; 1341 BP.
XX AAC50382;
AC
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 64633.
XX
KW Hybridisation assay; Genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

```
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142085.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0148171.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 3.6%; Score 183.2; DB 21; Length 1341;
Best Local Similarity 98.4%; Pred. No. 7.8e-22;
Matches 185; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2919 AGGGAGTAGAAGTCTCGTACCAATGGAATCCGAATATAGTACCTCCAGAGGTAC 2978
Db 482 AAGGAGTAGAAGTCTCGTACCAATGGAATCCGAATATAGTACCTCCAGAGGTAC 541
Qy 2979 TTAGTGACAGGGTTACGATGGTTCCAGCAGCTGATATTGGTCTTGGCGGGTTATTCITT 3038
Db 542 TTAGTGACAGGGTTACGATGGTTCCAGCAGCTGATATTGGTCTTGGCGGGTTATTCITT 601
```

QY 3039 TCGTTATATGGCTGGATATTTTACCTTTTCCGAGACGGATCTTCAGGGTGTGTACAGAA 3098
 DB 602 TCGTTATATGGCTGGATATTTTACCTTTTCCGAGACGGATCTTCAGGGTGTGTACAGAA 661
 QY 3099 AAGTAAGT 3106
 DB 662 AAATAAAT 669
 RESULT 8
 AAF22306
 ID AAF22306 standard; DNA; 163319 BP.
 XX AC
 AC AAF22306;
 XX
 DT 20-MAR-2001 (first entry)
 XX
 XX Arabidopsis thaliana chromosome 4 centromere.
 XX
 XX Centromere; microsome; vector; ds.
 XX
 XX Arabidopsis thaliana.
 XX OS
 XX Arabidopsis thaliana.
 XX
 XX WO200055325-A2.
 XX
 XX 21-SEP-2000.
 XX
 XX 17-MAR-2000; 2000WO-US07392.
 XX
 XX 18-MAR-1999; 99US-0125219.
 XX PR 01-APR-1999; 99US-0127409.
 XX PR 18-MAY-1999; 99US-0134770.
 XX PR 13-SEP-1999; 99US-0153584.
 XX PR 17-SEP-1999; 99US-0154603.
 XX
 XX (UYCH-) UNIV CHICAGO.
 XX
 XX Preuss D, Copenhaver G, Keith K;
 XX WPI; 2000-587529/55.
 XX
 XX Recombinant DNA construct comprising a plant centromere, useful for
 XX producing stably inherited microsome which can serve as vectors for
 XX the construction of transgenic plant and animal cells
 XX
 XX Claim 79; Page 1389-1451; 1449pp; English.
 XX
 XX The present invention relates to a recombinant DNA construct of a plant
 XX (Arabidopsis thaliana) centromere. The constructs are useful for
 XX producing stably inherited microsome which can serve as vectors for
 XX the construction of transgenic plant and animal cells expressing
 XX selected proteins such as hormones, enzymes, interleukins, clotting
 XX factors, cytokines, antibodies, and growth factors.
 XX
 XX Sequence 163319 BP; 53475 A; 30117 C; 28735 G; 50992 T; 0 other;
 XX
 XX Query Match 3.2%; Score 163.6; DB 21; Length 163319;
 XX Best Local Similarity 53.7%; Pred. No. 2.2e-18;
 XX Matches 404; Conservative 0; Mismatches 319; Indels 29; Gaps 2;
 XX
 QY 165 TTTACTTTTCTCAATTTTCATGCTGTGTATATTTAAGAAATTTGGAATAATATTAATTTT 224
 DB 120103 TTTAGTGTATAATAATATACGATGGTGGTGAAGCCATTTTGAAAAAGATTTTGTAT 120162
 QY 225 ATTAAATGGTTAGATGTTGTGTTAACTTAATGGCATTTGGTTAGATTTTGTAT 284
 DB 120163 ATTCAATCGGTTTGAATAT-TTTGTGAACCTCAAAATTTGATATCGGTTTAGAGATTTAAGT 120221
 QY 285 TGGTTTATTCATTTTCAAAATTTGATGGTTAAATTTGGATAAACCATGGAACCACTTA 344
 DB 120222 TGTTTAATTCATTTTACCAAAATTTGGTGGTT----- 120255
 QY 345 ACCATTACAACCTAATCTATTTTACTATCAAAACCAATTTGATCATCAATCATTTGACT 404

DB 120256 --CATTCGCAACCCCACTTAATTTGATCAATATAACCACTTTGACTCAATACTTATTTGAAC 120313
 QY 405 CATCAACTCATTTGAGTCAGCAAAATTTTCAACTCATTAGGGTTCATGGTTCAGTTGAGTTG 464
 DB 120314 CATCAACTCATTTGAGTCAGCAAAATTTTCAACTAATTTAGATTTCATGGATTGAGTTGAATTG 120373
 QY 465 AGTTGAGCCCATGAATTTTGACCCCATTTTGACACCCCTACATATGATCATAAAGTTAATAAT 524
 DB 120374 GGTTCATCCATGAACCTTAACCACTTTTGACACCCCTAATAAGCTATAAAAAACAAT 120433
 QY 525 CAAAAATTAATTAATTAATTAATTTTAAAGGAATTTTGTAGATTTCATTTAGCTTAATTT 584
 DB 120434 TCAAAATAGAGTAAGTGTATTTTCCCAATGTGACCTTTTATTTATAGATATCTTATTAAC 120493
 QY 585 TATATAAAAAACATAATGCAAAAGTACTAATGTATAGTTATTTTATTTTATTTTAAATATGC 644
 DB 120494 TGATAAACCACTGCACCTATATATGAGGACCAACAGTTATTAATAATTTTAAATGATTG 120553
 QY 645 AAAAATTAATTAATTAATTTTCACTTTTCACTTTTCACTTTTCACTTTTCACTTTTCACTTT 704
 DB 120554 ACGATAATGAAGTGTCTTATTTTATATTTTATATATGTTTACATAAACGATAGTATTAATGATTA 120613
 QY 705 CACCAACACCTACATTTTCATATTTCTCTTCACATTTTCATTTTCTTCTGCTCTTATTC 764
 DB 120614 TCAAGATTAGGTAATTTTATTTTATGTAATTTCTTAATGTTTATGTTTATTTATGCTCTCT 120673
 QY 765 TTCTTATCATCTTCATTTGCCCAATTTTTCATTTGACATTTGTCATCGTTACTTTGTTATGTTATG 824
 DB 120674 ATATATATATATATATATATCTTAATAGATACACAAAATGTTAAGTATTTTCCCTTTGTAT 120733
 QY 825 ATTTTGAACATTTTAATGTGAATCCCAACCGACTATAGGACTATCAAGAAGTTTCAAAC 884
 DB 120734 AGATTTGTATATTTTAAATGCAATATATTAATFCTTAGTTTGTTTTAAATAATTAATGAGTAA 120793
 QY 885 TTTTAAAAAAGATCTTGAACCTTTTAAAGCAA 916
 DB 120794 ATTAAAAAGCAAAATTAATTTGGAGAAAAAA 120825

RESULT 9

ABZ17294/c

ID ABZ17294 standard; DNA; 2000 BP.

 XX AC
 XX ABZ17294;

DT 21-JAN-2003 (first entry)

 XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 5099.
 XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

 XX OS Arabidopsis thaliana.
 XX PN WO200216655-A2.

 XX PD 28-FEB-2002.
 XX PF 24-AUG-2001; 2001WO-US26685.

 XX PR 24-AUG-2000; 2000US-227866P.
 XX PR 26-JAN-2001; 2001US-264647P.

 XX PR 22-JUN-2001; 2001US-300111P.
 XX

 XX PA (SCRI) SCRIPPS RES INST.
 XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

 XX PI Harper JF, Kreps J, Wang X, Zhu T;
 XX WPI; 2002-304127/34.

 XX DR
 XX PT Identifying a stress condition to which a plant cell has been exposed

XX and producing plants with increased tolerance to these abiotic stresses

PT -
 XX Claim 144; SEQ ID NO 5099; 577bp + Sequence Listing; English.
 PS
 XX The invention relates to identifying a stress condition to which a plant cell has been exposed, comprising:
 CC (a) contacting nucleic acid representative of expressed polynucleotides in the plant cell with an array or probes representative of the plant cell genome; and
 CC (b) detecting a profile of expressed polynucleotides in the plant cell characteristic of a stress response. The method is useful in the production of transgenic plants, cells and seeds and in producing plants with increased tolerance to abiotic stress. The present sequence is that of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used in methods of the invention.
 CC Note: The sequence data for this patent is not represented in the printed CC specification but is based on sequence information supplied to Derwent by CC the European Patent Office.
 XX
 SQ Sequence 2000 BP; 678 A; 314 C; 294 G; 714 T; 0 other;
 Query Match 3.0%; Score 152.4; DB 24; Length 2000;
 Best Local Similarity 91.0%; Pred. No. 1.2e-16;
 Matches 162; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
 QY 338 ACCATTAAACGATTACAACTTAACCTCACTTTTACTCATCAACCACTTGACTCACTCA 397
 DB 1557 ACCATTAAACGATTACAACTTAACCTCACTTTTACTCATCAACCACTTGACTCACTCA 1498
 QY 398 TTGACTCATCACTCACTTTGAGTCAAAATTTCACTCATTAGGTTTCATGGTTGAGT 457
 DB 1497 TTGACCATCACTCACTTTGAGTCAAAATTTGAACTTATGAGTTCTGGTTGAGT 1438
 QY 458 TGAGTTGAGTTGACCCGAAATTTTGACCCCATTTTGACCCCATTTGACATATGATCA 515
 DB 1437 TGAATTGAGTTGACCCGAAATTTTGACCCCATTTTGACCCCATTTGACATATGATCA 1380
 RESULT 10
 AAL48584/c
 ID AAL48584 standard; DNA; 1502 BP.
 AC AAL48584;
 DT 11-OCT-2002 (first entry)
 DE A thaliana WRKY64 promoter.
 KW WRKY; transgenic; plant; pathogen resistance; nematocide; fungicide;
 KW antibacterial; virucide; insecticide; promoter; ds.
 XX Arabidopsis thaliana.
 OS
 XX DE10063986-A1.
 XX 27-JUN-2002.
 XX 21-DEC-2000; 2000DE-1063986.
 XX 21-DEC-2000; 2000DE-1063986.
 XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA Lippok B, Somsich I, Eulgem T;
 PI WPI; 2002-576327/62.
 DR
 XX New transgenic plants with increased specific and non-specific
 PT resistance to pathogens, have increased expression or activity of
 PT specified nucleic acids -
 XX
 PS Claim 10; Page 28-29; 40pp; German.
 CC

CC The present invention relates to transgenic plants with improved
 CC resistance to pathogens as a result of increased expression of one or
 CC more of the nucleic acid sequences shown in AAL48564-AAL48574, or to the
 CC altered biological activity of these. The sequences can be used to
 CC improve plant resistance to a wide variety of nematodes, fungi, bacteria,
 CC viruses and animal pests (insects). The promoters from these genes can be
 CC used to drive expression of other genes. The present sequence is an
 CC Arabidopsis thaliana WRKY gene promoter described in the exemplification
 CC of the invention.
 XX
 SQ Sequence 1502 BP; 564 A; 257 C; 180 G; 501 T; 0 other;
 Query Match 2.3%; Score 115.8; DB 24; Length 1502;
 Best Local Similarity 77.9%; Pred. No. 1.7e-10;
 Matches 152; Conservative 0; Mismatches 42; Indels 1; Gaps 1;
 QY 1 TTTTGGCGGGAATCTCGGTTTACGTTTTCGCGGGAATCTCGGTTTACGTTT 60
 DB 196 TTTTGGCGGGAATCTCGGTTTACGTTTTCGCGGGAATCTCGGTTTACGTTT 137
 QY 61 TGGCGGGAATCTCGGTTTACGTTTTCGCGGGAATCTCGGTTTACGTTT 120
 DB 136 TGGCGGGAATCTCGGTTTACGTTTTCGCGGGAATCTCGGTTTACGTTT 78
 QY 121 CGGGAATCTCGGTTTACGTTTTCGCGGGAATCTCGGTTTACGTTT 180
 DB 77 CGGGAATCTCGGTTTACGTTTTCGCGGGAATCTCGGTTTACGTTT 18
 QY 181 TTCATTGCTTGATA 195
 DB 17 AAAATTACGGTTTA 3
 RESULT 11
 AAF22305/c
 ID AAF22305 standard; DNA; 1082138 BP.
 XX
 AC AAF22305;
 DT 20-MAR-2001 (first entry)
 DE Arabidopsis thaliana chromosome 4 centromere.
 XX Centromere; micromosome; vector; ds.
 KW
 XX Arabidopsis thaliana.
 OS
 XX WO200055325-A2.
 XX 21-SEP-2000.
 XX 17-MAR-2000; 2000WO-US07392.
 XX 18-MAR-1999; 99US-0125219.
 XX 01-APR-1999; 99US-0127409.
 XX 18-MAY-1999; 99US-0134770.
 XX 13-SEP-1999; 99US-0153584.
 XX 17-SEP-1999; 99US-0154603.
 XX (UYCH-) UNIV CHICAGO.
 XX
 XX Preuss D, Copenhaver G, Keith K;
 PI WPI; 2000-587529/55.
 DR
 XX Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsomes which can serve as vectors for
 PT the construction of transgenic plant and animal cells -
 XX
 PS Claim 68; Page 977-1388; 1449pp; English.
 XX
 CC The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for

CC producing stably inherited microsome which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.
 XX

SQ Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;
 Query Match 2.2%; Score 112.4; DB 21; Length 1082138;
 Best Local Similarity 81.2%; Pred. No. 1e-09;
 Matches 143; Conservative 0; Mismatches 31; Indels 2; Gaps 1;
 1 TTTTGGCGGAGAAATCTCGGTTTACGTTTGGCGGAGAAATCTCGGTTTACGTTT 60
 QY TTTATGTPAGAAAATTAACGGGTTTATGTTTGGCAAGCTAATAGGGGTTTACGTTT 678694
 Db TTTATGTPAGAAAATTAACGGGTTTATGTTTGGCAAGCTAATAGGGGTTTACGTTT 678694
 QY 61 TGGCGGAGAAATCTCGGTTTACGTTTGGCGGAGAAATCAACGGGTTTACGTTT 118
 Db TGGCGGAGAAATCTCGGTTTACGTTTGGCGGAGAAATCAACGGGTTTACGTTT 678694
 QY 119 TGGCGGAGAAATCAACGGGTTTATGTTTGGCGGAGAAATCAACGGGTTTACGTTT 174
 Db TGGCGGAGAAATCAACGGGTTTATGTTTGGCGGAGAAATCAACGGGTTTACGTTT 678568

RESULT 12
 ABZ10246
 ID ABZ10246 standard; DNA; 8056 BP.
 XX
 AC ABZ10246;
 XX
 DT 16-JAN-2003 (first entry)
 XX

DE Haematopoietic cell proliferation disorder related DNA sequence #386.
 XX
 KW Human; haematopoietic cell proliferation disorder; cytostatic;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200277272-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-EP03401.
 XX
 PR 26-MAR-2001; 2001US-278333P.
 XX

PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
 PI Pellet C, Schwope I, Ziebarth H;
 XX
 XX WPI; 2003-018942/01.
 DR

XX Detecting and differentiating between hematopoietic cell proliferative
 PT disorders, comprises contacting a target nucleic acid with a reagent
 PT that distinguishes between methylated and non-methylated CpG
 PT dinucleotides -
 XX
 PS Claim 28; SEQ ID 386; 117pp; English.
 XX

CC The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. AB209861 to AB211118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for

CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute for
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related
 CC DNA sequences. The nucleotide sequences from the present invention can
 CC also be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables
 CC a highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients.
 XX

SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;
 Query Match 2.0%; Score 105; DB 25; Length 8056;
 Best Local Similarity 45.5%; Pred. No. 1.3e-08;
 Matches 584; Conservative 0; Mismatches 675; Indels 24; Gaps 5;
 165 TTTACTTTTCTCAATTTTCATTTGTTGATATTTTAAAGAAATTTGAAAAATTTAATTTT 224
 QY TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1150
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1150
 QY 225 ATTAAATTTGTTTAAATGTTTAAATGTTTAAATTTGTTTAAATTTTTCGATTTT 284
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1210
 QY 285 TGGTTTAAATTTTAAATGTTTAAATGTTTAAATTTGTTTAAATTTTTCGATTTT 344
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1270
 QY 345 ACCATTACAACTAACCTATTTTCTCATCAAACTAACCTATTTTCTCATCACTTTGACT 404
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1330
 QY 405 CATCACTCAATTTGAGTCAAAATTTTCACTCACTCACTCACTCACTCACTCACTCACT 464
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1389
 QY 465 AGTTGACCCATGAATTTTGACCCATTTTGACCCATTTTGACCCATTTTGACCCATTTGAC 524
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1449
 QY 525 CAAAAATTTACTATTTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 584
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1509
 QY 585 TATATAAAAAACATAATGCAAAAGTACTAATGTATAGTTATTTTATTTTATTTTATATGC 644
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1569
 QY 645 AAAATATTTACTGTAATTTTCTCATTTCTTATCAATTTTCTTATTTTCAATTTAGC 704
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1628
 QY 705 CACCAACCTCATTTTCTTCTCATTTTCTTCTCATTTTCTTCTCATTTTCTTCTTCTTC 764
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1688
 QY 765 TCTTATCATCTCATTTGCCAAATTTTTCATTTGATTTGATTTGATTTGATTTGATTTG 824
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1748
 QY 825 ATTTTGTGAACATTTTAAATGTTGAAATTTTAAATGTTGAAATTTTAAATGTTTCAAC 884
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1808
 QY 885 TTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTTTAAATGTT 944
 Db TTTTAAATGTTTAAATGTTTATATTTGTTTAAATTTTTCGATTTTCAAAATTT 1868
 QY 945 TCTTCTTGTGTGATGAAGCTCTTCTGTCGCGGCGGAAATCTAGGCCATATAAGCCCTCTT 1904

peptide, for manufacturing complex carbohydrates, or as targets for
 screening GM4,6D antagonists for treating e.g. arthritis, or transplant
 rejection -
 XX
 XX
 Disclosure; SEQ ID NO 3933; 6pp: English.
 XX
 XX
 The invention relates to a composition comprising a human GDP-mannose
 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying
 GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation
 in a mammalian subject and for treating or ameliorating diseases affected
 by the level of cellular fucosylation or diseases affected by the
 fucosylation of glycoconjugates. These diseases include arthritis,
 transplant rejection, asthma, sepsis, reperfusion injury, stroke or
 infection. The GM4,6D peptide or a polynucleotide encoding it is also
 useful for manufacturing complex carbohydrates and as targets for
 screening small molecule antagonists of the activity of the enzyme. The
 polynucleotide is useful in developing an assay for defects in the
 enzyme, as well as in gene replacement therapy. Sequences
 ABX17942-BEX17944 and ABX17947-ABX33716 represent DNA molecules encoding
 human GM4,6D peptides of the invention.
 CC
 Note: The sequence data for this patent did not form part of the printed
 specification but was obtained in electronic format directly from USPTO
 at seqdata.uspto.gov/sequence.html.
 CC
 XX
 Sequence 404 BP: 113 A; 82 C; 87 G; 122 T; 0 other;
 XX

1349	ATACATATATGCTTATCTTTTAAATAAATTTACTAAATGATAGTGAATTATAGATTACT	1408
2285	TTTAAATTAATTATTTT-AAAATAAATTTTAAATTTTCTGATTTAAATAATTTTAT	2342
1409	TTTTTTTTTCTTTTTTAAATTTT	1431
2343	TTAATTTTTTAAATATATATTTTT	2365

RESULT 13
 ABX21874
 ID ABX21874 standard; cDNA; 404 BP.
 XX
 ABX21874;
 AC
 XX
 DT
 DT
 10-FEB-2003 (first entry)
 XX
 XX
 DE Human GMP-mannose 4,6-dehydratase (GMA_6D) DNA #3931

RESULT 14	
ABX62736	
ID	ABX62736 standard; DNA; 397 BP.
XX	
AC	ABX62736;
XX	
DT	25-FEB-2003 (first entry)
XX	
DE	Arabidopsis thaliana expressed sequence related polynucleotide #851.
XX	
KW	Transgenic plant; plant; genetically modified cell;
KW	environmental stress; ribozyme creation; disease resistance;
KW	stress tolerance; fungicide screening; insecticide screening;
KW	gene; ds.
XX	
OS	Arabidopsis thaliana.
XX	
PN	US2002040490-A1.
XX	
PD	04-APR-2002.
XX	
PF	26-JAN-2001; 200IUS-0770423.
XX	
PR	27-JAN-2000; 2000US-178512P.
XX	
PA	(GORL/) GORLACH J.

(GORL/) GORLACH J.

PA (ANY/) AN Y.
 PA (HAMI/) HAMILTON C M.
 PA (PRIC/) PRICE J L.
 PA (RAIN/) RAINES T M.
 PA (YUYI/) YU Y.
 PA (RAME/) RAMEKA J G.
 PA (PAGE/) PAGE A.
 PA (MATH/) MATHAW A V.
 PA (LEDF/) LEDFORD B L.
 PA (WOES/) WOESSNER J P.
 PA (HAAS/) HAAS W D.
 PA (GARC/) GARCIA C A.
 PA (KRICK/) KRICKER M.
 PA (SLAT/) SLATER T.
 PA (DAVI/) DAVIS K R.
 PA (ALLE/) ALLEN K.
 PA (HOFF/) HOFFMAN N.
 PA (HURB/) HURBAN P.
 XX
 PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
 PI Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
 PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
 PI Hurban P;
 XX
 DR WPI: 2003-110411/10.
 XX
 XX Novel Arabidopsis thaliana nucleic acid useful for identifying
 PT homologous or related genes, and to create genetically modified and
 PT transgenic organisms, such as plant cells and plants -
 XX
 XX Claim 1; SEQ ID NO 851; 43pp; English.
 XX
 CC The invention describes an Arabidopsis thaliana nucleic acid (I). The
 CC polypeptide (II) encoded by (I), transgenic plant (III) or genetically
 CC modified cell (IV) are useful for screening a candidate agent for its
 CC biological effect, by combining the candidate agent with (II), (III) or
 CC (IV), and determining the effect of the candidate agent on (II), (III) or
 CC (IV). (I) is useful for identifying homologous or related genes, for
 CC producing compositions that modulate the expression or function of its
 CC encoded protein, for mapping functional regions of the protein, in
 CC diagnosis, for studying associated physiological pathways, for genetic
 CC manipulation of cells, preferably plant cells, in screening assays of
 CC various plant strains to determine the strains that are capable of
 CC withstanding a particular disease or environmental stress, for enhancing
 CC or inhibiting production of biosynthetic product in a plant, for
 CC producing polypeptides, as probes for the detection of mRNA in biological
 CC samples, to generate additional copies of (I), to generate ribozymes or
 CC oligonucleotides, as single stranded DNA probes or as triple-strand
 CC forming oligonucleotides, and to create genetically modified and
 CC transgenic organisms, such as plant cells and plants. (II) or (III) is
 CC useful for introducing or improving disease resistance and stress
 CC tolerance in plants, screening biological active agents, e.g.,
 CC fungicides, insecticides, etc., and for elucidating biochemical pathways.
 CC (III) is useful as crops for their enhanced disease resistance, enhanced
 CC traits of interest, for screening programs, as crops which exhibit
 CC enhanced tolerance to environmental stress, or to produce a factor. This
 CC sequence represents a nucleic acid that may correspond to naturally
 CC occurring Arabidopsis thaliana expressed sequences.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC US patent office at
 CC seqdata.uspto.gov/sequence.html?docID=999909770423.
 XX
 SQ Sequence 397 BP; 114 A; 81 C; 81 G; 121 T; 0 other;
 Query Match 1.9%; Score 97.2; DB 25; Length 397;
 Best Local Similarity 69.5%; Pred. No. 2e-07;
 Matches 132; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
 QY 2910 TACTCAATCAGGAGTAGAAGTCTTCGCGTACCATGTGGAACTCCGAGACTATGTAGCTC 2969
 DB 28 TACCTGACACAGGAGTTACCACTCTTAAGACACAACTGTGGAACTCCCAATTACGTGCTC 87

QY 2970 CAGAGGTACTTAGTGGACAGGGTTACGATGTTTCAGCAGCTGATTAATTTGGTCTTGCGGGG 3029
 DB 88 CTGAGGTCTTCAGTCACAGGGTTACAAATGGTCCGTTGCAGATAATTTGGTCTTGCGGGG 147
 QY 3030 TTATTCCTTTTCGTTATATATTTGGCTGGATATTTACCTTTTTCGAGACGGATCTTCAGGGT 3089
 DB 148 TCATCCCTTTATGTTCTTTATGGCAGGATATCTTCCATTTGATGAATGGATCTACCAACTT 207
 QY 3090 TGTACAGAAA 3099
 DB 208 TATATAGTAA 217
 RESULT 15
 AAC42994
 ID AAC42994 standard; DNA; 1338 BP.
 XX
 AC AAC42994;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37631.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EPI033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143354.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.9%; Score 97.2; DB 21; Length 1338;
Best Local Similarity 69.5%; Pred. No. 2.2e-07;
Matches 132; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy	2910	TACTCAATCAGGAGTAGA	2969
Db	467	TACCTGACACAGGATTACCATCCTAAGACACATGTGGA	526
Qy	2970	CAGAGGTACTTAGTGGACAGGGTTACGATCGTTACGACGTGATATTTGGTCTTGGGGG	3029
Db	527	CTGAGGTTCTCAGTTCACAGGGTTACAAATGGTGGCGTTGCAGATATTTGGTCCCTGTGGGG	586
Qy	3030	TTATTCCTTTTCGTTATATTGGCTGGATATTACCTTTTCCGAGACGGATCTTCCAGGTT	3089
Db	587	TCATCCCTTATGTTCTTATGGCAGGATATCTTCCATTTGATGAATGGATCTACCACTT	646
Qy	3090	TGTACAGAAA	3099
Db	647	TATATAGTAA	656

Search completed: October 21, 2003, 01:44:08
Job time : 1224 secs